DOCTOR OF PHILOSOPHY

Study of the effect of DNA-Methylation on meiotic recombination in Arabidopsis thaliana and barley (Hordeum vulgare)

Sourdille, Adeline

Award date:
2021

Study of the effect of DNA-Methylation on meiotic recombination in *Arabidopsis thaliana* and barley (*Hordeum vulgare*)

Adeline Sourdille

Thesis presented in fulfilment of the requirement for the degree of

Doctor of Philosophy

University of Dundee

2016-2021

Abstract

Meiotic recombination underpins both applied plant breeding and gene mapping in fundamental research. However, in large-genome crops such as cereals (including barley; *Hordeum vulgare*), recombination generally occurs close to the telomeres with around 30% of the genes rarely, if ever, recombining. Understanding recombination in cereals is therefore crucial. So far, most meiotic research in plants has focused on understanding mechanistic aspects of the formation of crossovers (CO), but little has been centred on the effect of epigenetic markers, including DNA-methylation, on recombination. However, manipulating the epigenome could have the potential to release novel combinations of genetic diversity.

Maintenance of pre-existing DNA-methylation is mainly driven by *Methyltransferase 1* (*Met1*), which is involved in CG-methylation. In *Arabidopsis thaliana*, when compared to wild type (WT), hypomethylated *met1* mutants exhibit a higher CO frequency at the ends of chromosomes and decreased levels in the peri-centromeric regions. However, the overall number of COs in the genome remains constant. Similar trends were also observed in *DNA-demethylation 1* (*ddm1*) mutants in *Arabidopsis*.

In this study, we aim to manipulate DNA-methylation in barley and *Arabidopsis*, both transiently and by mutagenesis. Zebularine, a demethylating cytidine analogue, was applied to *Arabidopsis* and barley F_1 hybrid seeds in an attempt to phenocopy *met1* mutants. In *Arabidopsis*, three Fluorescent Tagged Lines (FTL) were used to visualise recombination between two markers directly in the seeds. These markers spanned peri-centromeric, interstitial and sub-telomeric loci on the chromosomes. Changes in recombination were observed in the sub-telomeric interval but not in the interstitial centromeric intervals. Additionally, gene expression was measured using RT-qPCR to assess the general effect of zebularine on plant development, showing significant changes in gene regulation in the presence of zebularine in genes involved in germination, vegetative/flowering stage balance, stress response and DNA-methylation. In barley, zebularine treatment triggered delayed development during germination but the plants quickly recovered, and no effect was observed on the recombination landscape when measured using SNP genotyping. Gene expression was also analysed using

microarray, but the effect of the zebularine treatment was too stochastic on the seedlings to generate conclusions on its effect on gene regulation.

In parallel, a large TILLING (Targeting Induced Local Lesions in Genomes) population was used to identify a collection of barley *met1* mutants (*cv.* Golden Promise). Two lines carrying missense mutations were identified where the nucleotide change is predicted to have a highly deleterious effect on protein function. These lines were crossed into another WT background (*cv.* Barke) and F_3 families were generated. The plants were characterized for effects on plant performance and were then genotyped using a 50k SNP genotyping array. F_3 genotyping shows a small tendency to redistributed recombination frequency in *met1* mutants in a non-significant manner. Finally, a CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats) construct was developed to target mutagenesis in *met1* and *ddm1* in barley *cv*. Golden Promise.

Acknowledgements

I would like to thank Pr. Robbie Waugh, and Dr. Luke Ramsay for directing me in my project whilst leaving me the necessary liberties to explore this fascinating side of research. You have been a bottomless well of knowledge and have helped me grow to become the scientist I am today.

This PhD project would not have been possible without the financial and moral support of the Gatsby foundation. Not only did this allow me to carry out the work in this thesis, but they introduced me to an incredibly supportive network of fellow students and alumni with who I look forward to continuing collaborating.

I would like to express my gratitude to all members, past and present, of the Barley Meiosis group at the James Hutton Institute for their support and good humour as well as the near-endless supply of cakes during our meetings. Especially, I would like to thank Dr. Mikel Arrieta for his invaluable help with genotyping analysis and Dr. Miriam Schreiber for her mentoring on all things bioinformatics.

My thanks also go to Pr. Ian Henderson and all the members of his lab in the University of Cambridge for introducing me to the world of *Arabidopsis* research and the magic of fluorescent seed lines. Even though I was only a part of the group very briefly, you have made me feel at home right away and I have learned a lot during these three months.

I am particularly grateful for the assistance provided by Dr. Monika Spiller (Syngenta) who helped providing me with F1 hybrid seeds (Jettoo) which were an invaluable asset in this PhD project.

I would also like to thank my colleagues of the Molecular Biology team at Tropic Biosciences for their support over the past six months. From day one, you have been able to encourage me into continuing to steadily write the final parts of this manuscript, despite Covid lockdowns and long days at work. In particular, my thanks to our Team Leader Dr. Angela Chaparro for being so flexible on letting me organise my time to finish writing this thesis in the best conditions.

My life as a PhD student would have been very dull without my fellow PhD students, friends, and colleagues, in and out of the James Hutton Institute. Thank you Akis, Euan,

Jenny, Jonny, Kerry, Shona, and Trisha. I do hope we will have many more opportunities to explore Scotland and maybe other parts of the world together.

I want to thank Patricia and Andrew Price for accepting me as an additional member of the family for the past four years and your never-ending support and help.

I owe my deepest gratitude to my parents, Sandra and Pierre Sourdille, as well as to my grandparents, Madeleine and Paul “Yote” Tixier, for letting your only daughter and granddaughter leave the French nest and fly to Scotland. You have been supporting every decision I made and have always provided me with insightful advice when doubt hit. I feel particularly privileged to have been working in the same field of research as my dad and will always treasure your help and our scientific exchanges, past and future.

Finally, I would like to thank my partner, James Price. Over the past four years you have been my rock, supporting me, cheering me up, pushing me, scolding me when I needed, and you will never know how much that has meant to me.

Declarations

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or processional qualification except as specified.

Adeline Sourdille

Publications

Schreiber, M., Barakate, A., Uzrek, N., Macaulay, M., **Sourdille, A.**, Morris, J., Hedley, P. E., Ramsay, L., & Waugh, R. (2019). A highly mutagenised barley (cv. Golden Promise) TILLING population coupled with strategies for screening-by-sequencing. *Plant Methods*, *15*(1). https://doi.org/10.1186/s13007-019-0486-9

Table of contents

Table of Figures

Table of tables

Abbreviations

BAH	Bromo-Adjacent Homology
BSMV	Barley Stripe Mosaic Virus
CO	Crossing Over/Crossover
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTL	Col-0 Traffic Line
cv.	Cultivar
DDM-1	Decreased DNA Methylation 1
dHJ	Double Holliday Junction
DMR	Differentially Methylated Region
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DSB	Double Strand Break
EMS	Ethyl Methanesulfonate
epiRIL	Epigenetic Recombinant Inbred Lines
FTL	Fluorescent Tagged Line
GBM	Gene Body Methylation
GBS	Genotyping By Sequencing
gDNA	Genomic DNA
GM	Genetically Modified
GWAS	Genome Wide Association Study
HDR	Homology Directed Recombination
HPLC	High Performance Liquid Chromatography
MET-1	Methyl Transferase 1

mRNA	Messenger RNA
MS	Mass Spectrometry
NCO	Non-Crossover
ncRNA	Non-Coding RNA
NGS	Next Generation Sequencing
NILS	Nearly Isogenic Lines
PCR	Polymerase Chain Reaction
PROVEAN	Protein Variation Effect Analyser
RDdM	RNA-Directed *de novo* Methylation
RISC	RNA-Induced Silencing Complex
RNA	Ribonucleic Acid
RNAi	RNA Interference
RNAseq	RNA Sequencing
RRBS	Reduced Representation Bisulfite Sequencing
RT-qPCR	Reverse Transcriptase Quantitative PCR
SC	Synaptonemal Complex
sgRNA	Single Guide RNA
siRNA	Small Interfering RNA
SNP	Single Nucleotide Polymorphism
TE	Transposable Element/Transposon
TILLING	Targeting Induced Local Legions IN Genomes
VIGS	Virus Induced Gene Silencing
WGBS	Whole Genome Bisulfite Sequencing
WT	Wildtype

1. Introduction

1.1. Barley in agriculture and research

1.1.1. History of barley

Barley (*Hordeum vulgare* ssp. *vulgare L.*) is a member of the Poaceae family, most commonly known as grasses. The first evidence of domesticated barley dates to around 10,000 years ago. Prior to this, its wild relative, *Hordeum spontaneum,* was being increasingly used by ancient farmers and over time, selection for traits more suited to what we know today as agriculture established a foundation for the barley cultivars that are being grown in farmers' fields today (von Bothmer & Jacobsen 2015). Alongside wheat and many other major cereals, barley was first cultivated in the Fertile Crescent in the area around Jordan and Israel.

Today, barley is mainly used for human consumption either through the production of alcoholic beverages such as beer, whisky, and as a staple for production of bread and porridges (Baik & Ullrich 2008). Over time, in many areas of the world, wheat-based products and potatoes progressively replaced foods made from barley, which is now only marginally used as a staple food for human consumption. Currently, barley is generally used for animal feed (60% of the global production) and the brewing and distilling industries (Ullrich 2010). However, given its resilience and adaptability to a wide variety of environments and growing conditions, it remains a staple food crop in certain regions such as Tibet and Ethiopia (Baik & Ullrich 2008, Mohammed *et al.* 2016).

Barley is recognised as one of the most genetically diverse cereals. Distinctions are made between winter and spring barley, whether the grains are hulled or hull-less, and whether the spikes are 2-rowed or 6-rowed (Knüpffer *et al.* 2003). Additionally, barley can be classified depending on its end use, animal feed or for malting. The diversity and versatility that characterise barley make it a major crop in the world and underline its economic importance as livestock feed and in the production of alcoholic drinks.

1.1.2. Economic importance of barley

As of 2018, barley is the 4th most produced cereal crop in the world behind maize, rice, and wheat (Figure 1.1.1.A). Production amounted to 141 million tons in the world in 2018, with an average price of $198/ton in Europe (FAO 2020,

http://www.fao.org/faostat/en/#data/QC/visualize). Europe is the main continental producer of barley, producing 59% of the global total (Figure 1.1.1.B). In 2018, with 6.51 million tons, the United Kingdom was the 9th largest barley producer in the world (Figure 1.1.1.C). In Europe over the past 50 years, barley yields have increased by 60% (Figure 1.1.1.D). This is primarily due to targeted breeding producing high yielding varieties, which are more disease and insect resistant in addition to better agricultural practices which allow higher yields from a reduced area. However, the global yields for cereal crops production do not keep up with the ever-increasing world population, and there is a constant need for the creation of new varieties with higher yields. In addition, there is an increasing demand for varieties that are more resilient to biotic and abiotic stresses and have improved sustainability in agriculture. Therefore, it is necessary for further barley breeding research to focus on identifying novel traits and unlocking diversity.

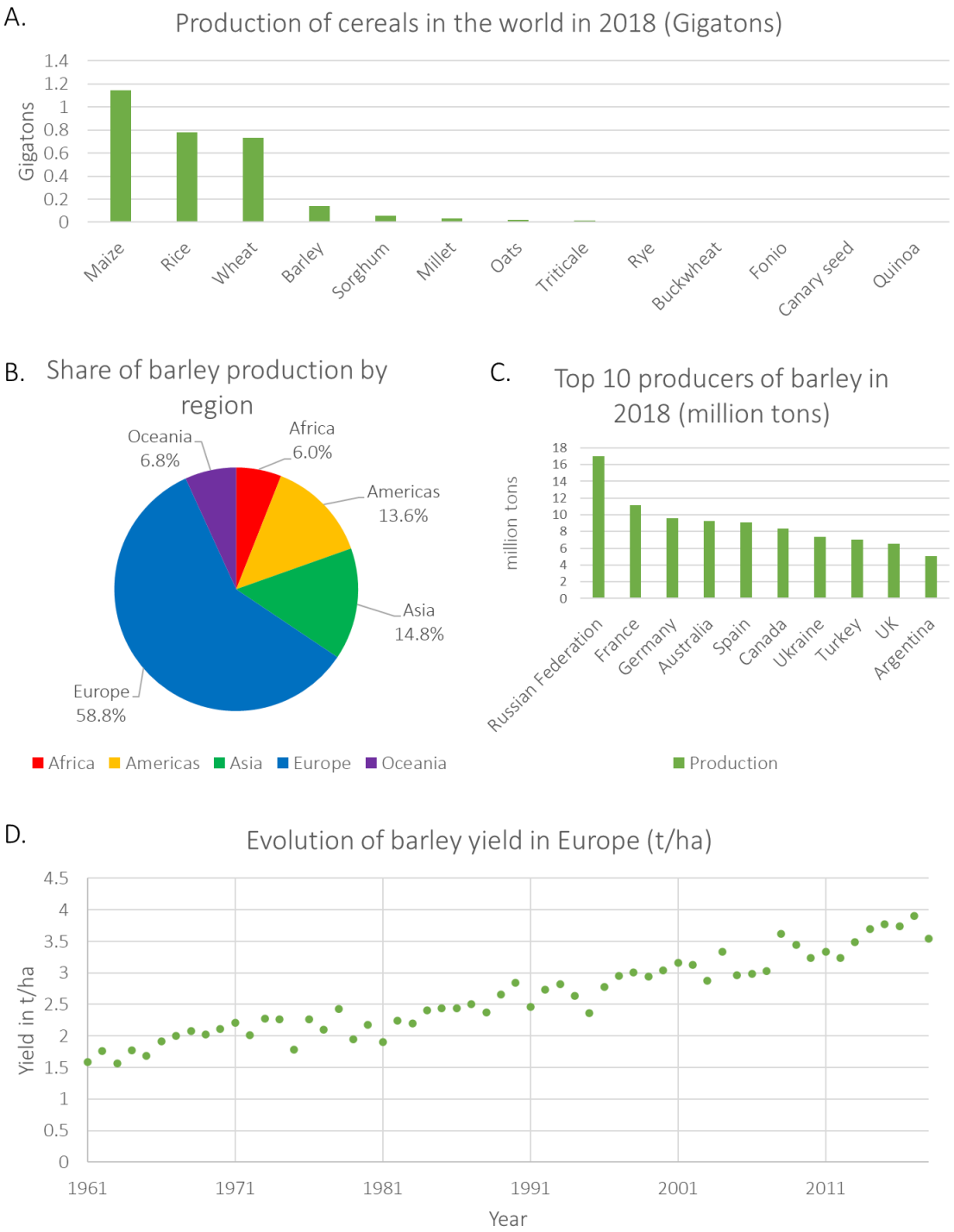


Figure 1.1.1. The economic importance of barley in the world and in the UK (data collected from FAOstat, 2020). A: Global production of the main crops in 2018 in Gigatons. Barley has the 4th production following maize, rice and wheat. B: Repartition of barley production in 2018 between the world regions. Europe dominates the market, producing almost 60% of the global production. C: Ranking of the top 10 barley producing countries in the world in 2018 shows the United Kingdom as the 9th barley producer in the world. D: Evolution of global barley production yield from 1961 to 2018 in Europe, in tons/hectares. Yields have improved by 60% in the last half-century.

1.1.3. Barley research

Barley is extensively used as a model for large genome cereals studies, especially for the Triticeae (subfamily Pooideae). As a well-studied species, barley benefits from its genome being organised into a low number of chromosomes and behaving as a true diploid species (2n = 2x = 14). Additionally, despite possessing a large genome of 5.3 Gbp (Mayer *et al.* 2012), this is relatively small compared to the 17 Gbp polyploid wheat genome (Choulet *et al.* 2014, Appels *et al.* 2018).

Extensive collections of genetically diverse barley genotypes (accessions) are present in gene banks around the world. For example, The John Innes Centre houses a large public barley collection in its Germplasm Resource Unit, which contains over 10,000 *Hordeum vulgare* accessions in addition to some 200 *Hordeum spontaneum* accessions (John Innes Centre 2020, https://www.jic.ac.uk/research-impact/germplasm-resource-unit/). These collections consist of seeds from a wide variety of elite varieties, landraces, and wild accessions, as well as several mutant collections (Druka *et al.* 2011, Comadran *et al.* 2011, Comadran *et al.* 2012). In particular, the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) in Gatersleben houses over 21,000 barley accessions which were genetically profiled using genotyping-by sequencing (GBS) (Milner *et al.* 2019). Large-scale analysis of this germplasm collection highlighted the need to add wild relatives to better access ancestral genetic diversity. Genome-Wide Association Studies (GWAS) on this population using common phenotypes such as number of rows in the floret (two-rowed vs. six-rowed varieties) and awn roughness identified two loci for each trait on chromosomes 2H and 4H and on chromosomes 5H and 7H, respectively (Milner *et al.* 2019). This further confirmed the potential of genotyping large populations to identify regions of allelic variation for genes of interest.

In recent years, barley research has benefitted from the development of several genomic resources. A draft whole genome sequence was made available in 2012, using *cv.* Morex as a reference genome (Mayer *et al.* 2012). This included a 4.98 Gbp physical map of the genome, of which 3.90 Gbp were anchored to a high-resolution genetic map. The sequence resource included extensive annotations such as gene names and descriptions, RNA sequences, transcripts, and their alternative splicing patterns etc. These resources were progressively completed over the following years, and in 2017 a chromosome-scale assembly of the barley genome was achieved through a large international collaborative

effort (Beier *et al.* 2017). A pseudomolecule was obtained for each barley chromosome (1H to 7H), with an additional pseudomolecule grouping all assembled sequences, which could not be anchored to the chromosomes (Un). This study identified over 39,000 genes in the barley genome that was overall made of 80.8% of transposable elements (TE). Generally, this research showed that the barley genome is characterised by a complex structure and large regions of DNA repeats.

Barley research further benefited from the development of a large genotyping array developed in 2017 containing 49,267 (50k) Single Nucleotide Polymorphisms (SNPs). This Barley 50k iSelect SNP Array constitutes a major tool in studying genetic diversity in barley cultivars for the breeding industry and for genetics research in academia (Bayer *et al.* 2017). Similarly, a microarray assay was developed using the transcripts described in the 2012 Morex assembly of the barley genome. This microarray allows rapid analysis of gene expression patterns of any barley cultivar without the need for RNAseq (Morris & Hedley 2019). Both these tools were developed at least in part by researchers from the James Hutton Institute and have been used during this PhD project for data analysis (Chapters 4 and 5).

Like many large-genome cereals, the distribution of crossing-over (COs) along the barley chromosomes is skewed towards the sub-telomeric ends of the chromosomes (Mayer *et al.* 2012). This means 20-30% of barley genes are locked in large linkage blocks that rarely, if ever, recombine (Baker *et al.* 2014). Physically, barley chromosomes generally form only a couple of chiasmata per chromosome in their distal ends and can be observed cytologically by the formation of seven ring-bivalent chromosome pairs during meiosis, at Metaphase I (Gale & Rees 1970). This is of major importance for the breeding industry and the academic community, as effectively only 50% of the genome can be used for breeding and genetic studies (Baker *et al.* 2014). Breeders and scientists are therefore strongly interested in ways of understanding and even controlling how and where recombination events are managed during meiosis.

Barley is an excellent model for understanding meiotic processes in large-genome cereals due to its diploid nature and large genetic diversity. However, its large genome and long generation time make it a difficult organism to use within a laboratory setting. This makes the use of a simpler model organism, such as *Arabidopsis thaliana*, with shorter

generation times and a smaller genome, of paramount importance both for gene discovery and for testing complex mechanisms.

1.2. Meiosis in plants

1.2.1. Overview of the meiotic process

Most eukaryote species rely on sexual reproduction to transmit their genetic materials and shuffle allelic combinations between generations. Sexual reproduction requires the generation of haploid cell lines in order to ensure correct ploidy in the resulting zygote after fertilisation. In this way, the obtained zygotes inherit half of their genetic material from each parent. The following description of the meiotic processes in plants has been adapted from a review by Mercier *et al.* (2015) and mostly involves genes discovered and characterised in *Arabidopsis*.

The creation of haploid gametes in plants is the result of a complex biological process occurring over 48 hours called meiosis. Meiosis consists of two steps of cell division, which generate four haploid cells from a single diploid cell. Similar to mitosis, each meiosis step is divided in four different stages, which can be cytologically differentiated from one another: prophase, metaphase, anaphase, and telophase. During the first meiotic division, called Meiosis I, pairs of homologous chromosomes are separated after an exchange of genetic material between them during meiotic recombination, resulting in two haploid cells (dyad stage). CO events are mandatory for the correct meiotic division and to produce viable gametes (Jones & Franklin 2006, Martini *et al.* 2006). In the same manner as mitosis, Meiosis II results in the segregation of sister chromatids to separate poles in the cells, resulting in the production of four haploid gametes (Figure 1.2.1).

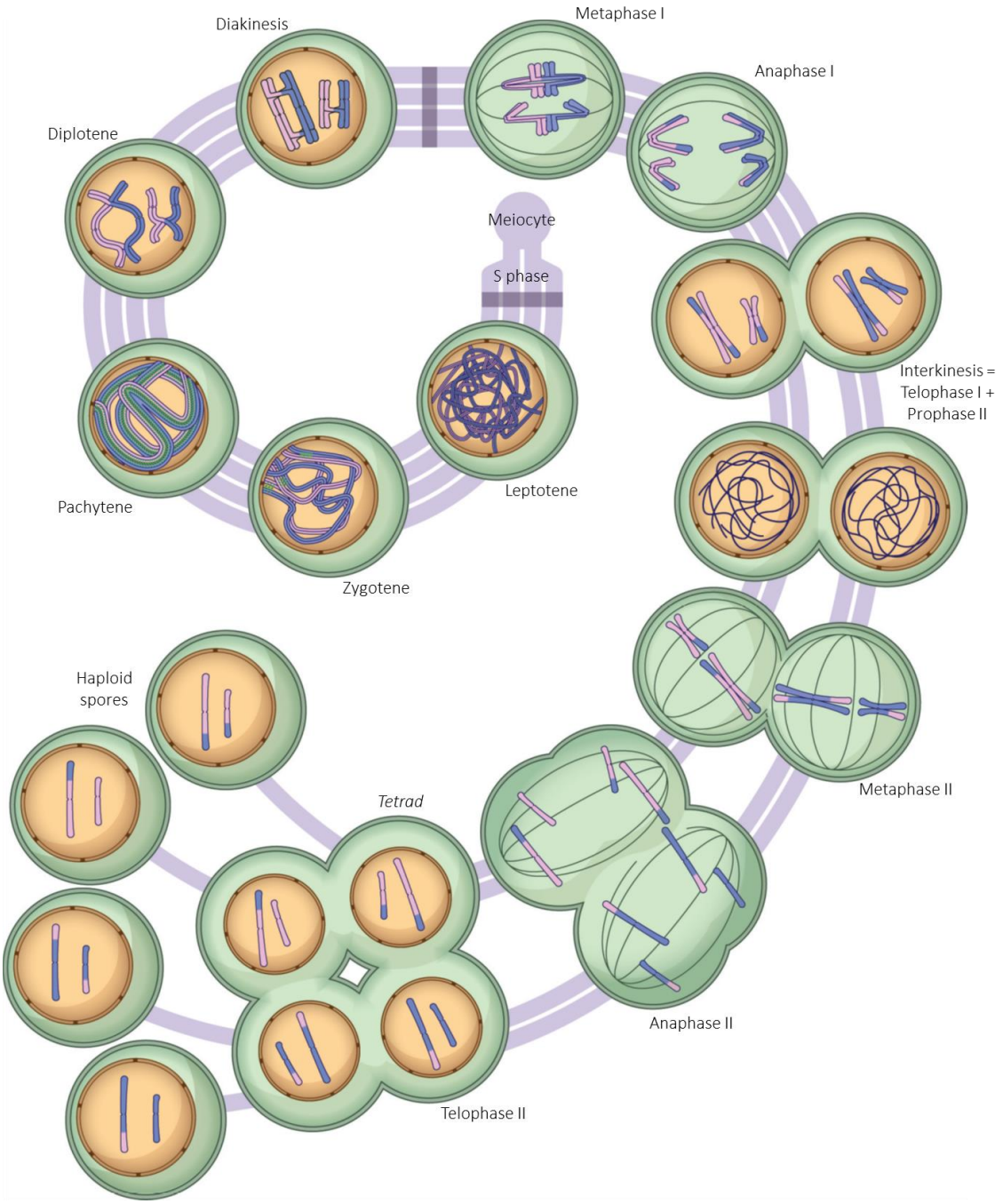


Figure 1.2.1. Overview of meiosis in plants. During the first meiotic division, chromosomes align, and CO formation starts during Prophase I, leading to the formation of bivalents during diplotene and dikinesis. At Anaphase I, the pairs of chromosomes separate and migrate to the poles, leading to for formation of two haploid nuclei containing bichromatids. During Meiosis II, similarly to mitosis, sister chromatids are separated and four haploid nuclei form, and the cell membranes eventually separate during cytokinesis. The resulting haploid spores will then mature into pollen. Adapted from Mercier *et al.* 2015.

It is important to note that in higher plants such as angiosperms, the development of the female gametophyte, the ovule, can vary significantly from the production of the male gametes, or pollen. For the female gametophyte, after the meiotic divisions, the obtained haploid cells usually undergo several rounds of mitosis, with only one of the haploid cells becoming the actual fertile gamete. The overall result is the formation of complex mature

female gametophytes which can be composed of several differentiated cells, some of which contain several nuclei (Armstrong & Jones 2001, Bennett *et al.* 1973, Yadegari & Drews 2004). The complexity of female meiosis in angiosperms added to the small number of fertile gametes being produced means that most studies of the meiotic process are done on male meiosis. Male meiosis occurs in the anthers, where originally undifferentiated cells called meiocytes undergo meiotic cell divisions that are subsequently followed by mitotic divisions to produce pollen in large quantities (Mercier *et al.* 2015). The processes discussed hereafter, as well as the following cytological analyses, have therefore been characterised in male meiocytes in plants. The fundamentals of many of these processes were previously discovered and characterised in other model eukaryotes such as budding yeast (*Saccharomyces cerevisiae*) and *Caenorhabditis elegans*.

1.2.2. Prophase I

Prophase I is widely recognised as perhaps the most important meiotic step (Ronceret & Pawlowski 2010), as it is during this stage that homologous recombination occurs between chromosomes, arising through the formation and repair of Double Strand Breaks (DSBs) (Keeney *et al.* 1997). This meiotic stage can be divided into five sub-stages: leptotene, zygotene, pachytene, diplotene and diakinesis. Cytological observations of these stages in barley and wheat (Colas *et al.* 2017) are illustrated on Figure 1.2.2.

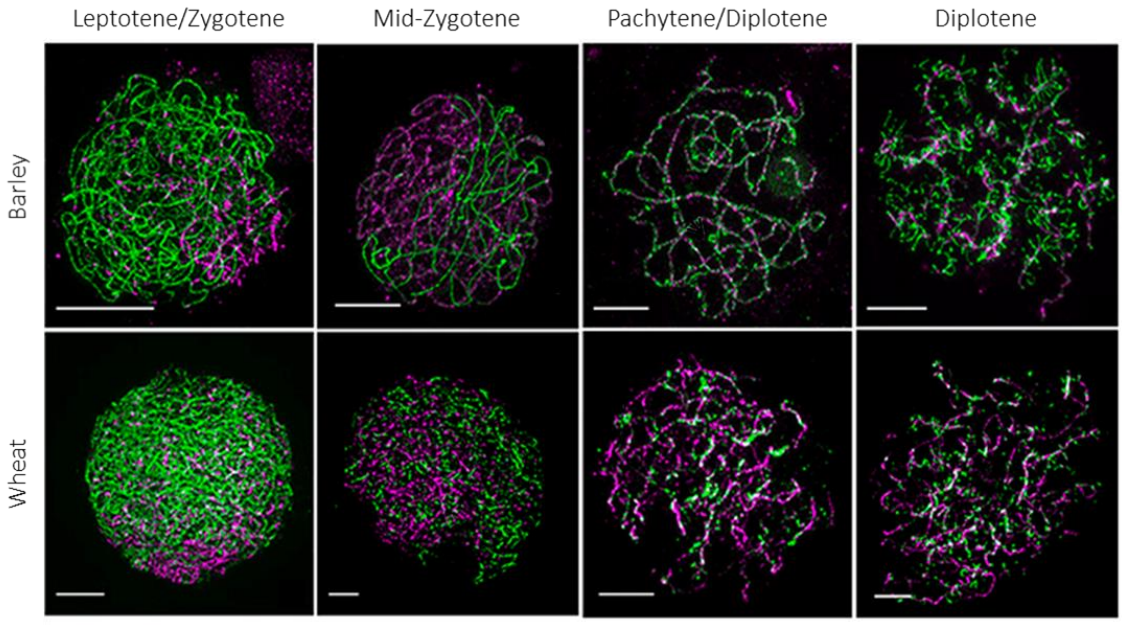


Figure 1.2.2. 3D-SIM microscopy of chromosome synapsis during Meiosis I in barley (*cv.* Bowman) and wheat (*cv.* Chinese Spring). The evolution of synapsis was followed using anti-ASY1 (green) and ZYP1 (magenta) antibodies. The roles of ASY1 and ZYP1 are explained later in this chapter. Scale bar = 5 µm

After an initial pre-meiotic DNA replication at G2 phase, meiocytes enter meiotic prophase at the leptotene stage. Chromatin is condensing, making the chromosomes easily observed under the microscope. This is due to the recruitment and formation of axial elements at the base of chromatin loops (Kleckner 2006). These elements will later also play a large role in the recruitment of the proteins, which constitute the synaptonemal complex and the alignment of homologous chromosomes during pairing. Cytologically, leptotene is characterised by the formation of a telomere bouquet (Bass 2003). Telomeres gather while attached to the nuclear envelope, facing the nucleolus. This phenomenon is thought to initiate the pairing of homologous chromosomes as telomeres are close to each other (Liebe *et al.* 2004).

The actual pairing of homologous chromosomes is initiated in zygotene. It is during this stage that the zipper-like synaptonemal complex (SC) is assembled between homologous chromosomes (Colas *et al.* 2008, Phillips *et al.* 2012, Higgins *et al.* 2012). This tripartite complex is composed of two lateral elements brought together in a zipper-like manner by a central element (White *et al.* 2004, Sym & Roeder 1995). More precisely, the meiotic protein HOP1 (also known as ASY1, Asynaptic 1, in plants) associates to the axial elements to help form the lateral elements, whereas the central element ZIP1 forms a dimer which brings both chromosomes together. Many other proteins are involved in this process, notably ZIP2 and ZIP4, which are believed to be involved in the recruitment of ZIP1 (Tsubouchi *et al.* 2006). Additionally, ZIP3 could be involved in regulating the SC formation by changing ZIP1 stability and therefore avoiding non-homologous recombination (Agarwal & Roeder 2000, MacQueen & Roeder 2009).

When the meiocytes enter pachytene, the SC formation is complete. Pachytene is the stage during which homologous recombination is thought to occur (Mercier *et al.* 2015). Cytologically, homologous recombination can be observed by the physical appearance of a chiasma for each CO event (Zakharyevich *et al.* 2012). With homologous chromosomes being linked usually by two distal chiasmata, so-called ring-bivalent chromosomes can be observed cytologically at the end of pachytene and beginning of diplotene (Wagenaar 1960). Homologous recombination and the formation of chiasmata will eventually ensure correct segregation of homologous chromosomes in Anaphase I (Martini *et al.* 2006, Jones & Franklin 2006).

During diplotene, the SC is disassembled. In barley, this dissolution has been shown to be highly organised, resulting in a characteristic tinsel chromosome conformation when observed immunocytologically under the confocal microscope (Colas *et al.* 2017). When the SC is fully disassembled during diakinesis, each pair of chromosomes will align on the equator of the cell, ready to enter Metaphase I (Osman *et al.* 2011).

1.2.3. End of Meiosis I and Meiosis II

During Metaphase I, the meiotic spindle is appearing, and the chromosomes are aligned on the metaphase plate (Osman *et al.* 2011). This, combined with the existence of chiasmata and cohesion complexes at the centromere of chromosomes, ensures only homologous chromosomes, and not the sister chromatids, are separated (Revenkova *et al.* 2004). The homologous chromosomes are then pulled towards their respective poles during Anaphase I (Koszul & Kleckner 2009) and the cells enter the second meiotic division immediately, without a DNA replication step, considering the chromosomes are still in a dyad conformation.

Similar to a mitotic division, during Prophase II, microtubules form a second spindle whilst chromosomes condense even further. The chromosomes then position themselves on the metaphase plate during Metaphase II, before the sister chromatids are separated and pulled in opposite directions toward the poles during Anaphase II. Finally, the nuclear envelope reforms around the newly separated single-chromatid chromosomes at Telophase II. In the case of pollen, cytokinesis then occurs, during which the cytoplasm is divided, which results in four separate haploid cells, or tetrads, which will then mature into pollen (Mercier *et al.* 2015).

1.2.4. Homologous recombination

Meiotic division and homologous recombination in sexually reproducing plants are two biological processes that are often tightly intertwined and co-dependant. The initiation of homologous chromosomes pairing is facilitated by the creation of DSBs, formed by the topoisomerase SPO11 (Keeney *et al.* 1997, Thacker *et al.* 2014, Bhuiyan & Schmekel 2004). These DSBs occur at very specific sites in the genome, though it is not yet fully understood how these sites are selected. Despite its role in chromosome pairing, SPO11 is not involved in the formation of the SC (Bhuiyan & Schmekel 2004). The formation of DSBs is then followed by resection of the separated DNA into single 3’ DNA overhangs by the MRN complex (MRE11, RAD50 and NBS1) (Gallego *et al.* 2001, Gao *et al.* 2009, Ji *et*

al. 2013). Meiotic proteins RAD51 and DMC1 then coat the overhang of single strand DNA and form a nucleoprotein filament (Brown *et al.* 2015). This will in turn facilitate the search for homologous sequences in the homologous chromosomes and trigger strand invasion and the creation of a joint molecule between homologous chromosomes (da Ines *et al.* 2012, Lorenz *et al.* 2014). This results in the formation of a D-loop that can then be resolved by the formation of a double Holliday Junction (dHJ), which almost can then result in the formation of a CO. If no dHJ is formed, the strand invasion will be aborted, resulting in a non-crossover (NCO) event (Allers & Lichten 2001, Patel *et al.* 2017). This process is summarised in Figure 1.2.3.

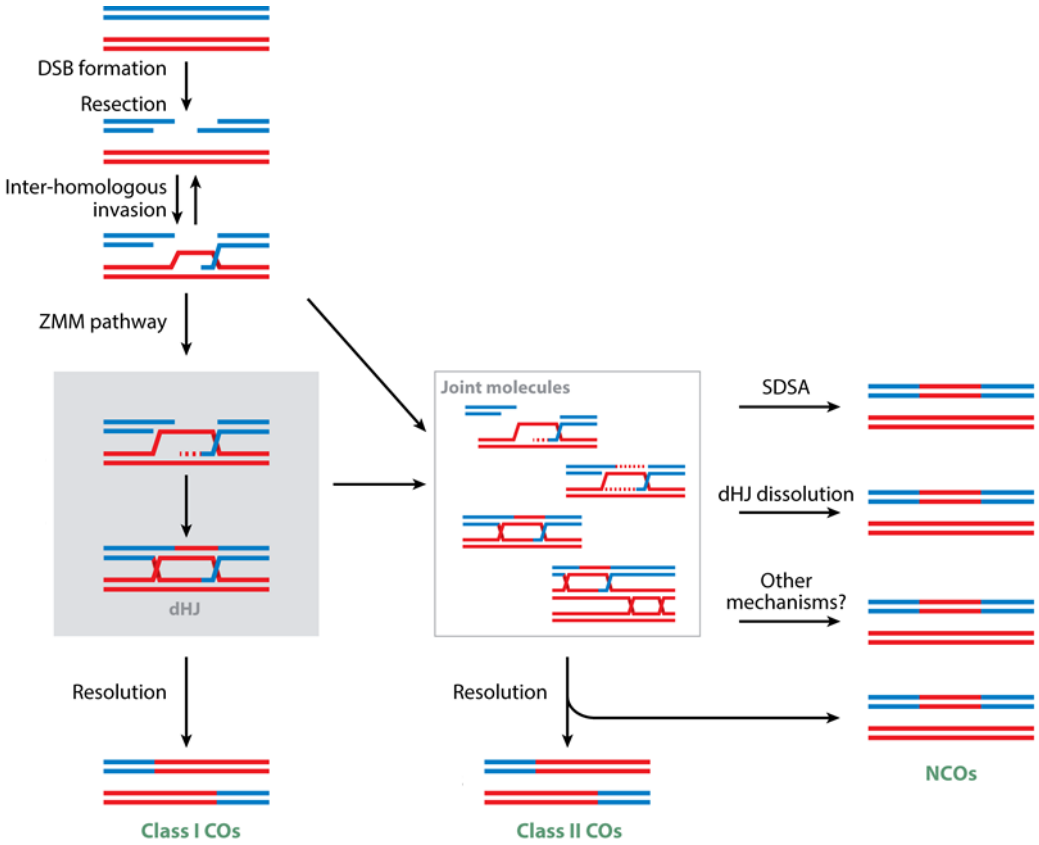


Figure 1.2.3. Simplified model of the meiotic recombination mechanisms in *Arabidopsis thaliana*. Adapted from Mercier *et al.* (2015)

In plants, only a small number of DSBs are resolved as CO events. For example, in *Arabidopsis*, cytological analysis of the number of RAD51 foci reveals the formation of over 150-200 DSBs (Xue *et al.* 2018), yet only an average of 10 chiasmata can be observed in the later stages (Higgins *et al.* 2004). The distribution of CO events is not random and seems to be limited to a small number of hotspots in the genome. (Mercier *et al.* 2015). This suggests the formation and location of CO events are regulated by a complex mechanism. While it is clear that the obligatory COs have a role in the correct segregation

of homologous chromosomes (Martini *et al.* 2006), the roles of the DSBs resolved as NCOs are still unclear. A commonly accepted hypothesis would be that these DBSs are crucial for appropriate chromosome alignment and the formation of the SC in the early stages of meiosis (Keeney *et al.* 2014).

Additionally, the formation of CO events in close vicinity to other crossovers is largely impaired by a complex mechanism termed CO interference that is yet to be fully elucidated. Understanding meiosis is made even more complex by the existence of a type of COs which are thought to be insensitive to CO interference (Higgins *et al.* 2004). Such COs, although in a very small proportion, have been called Class II COs, in contrast to interference-sensitive COs, labelled as Class I COs.

A recurring problem in meiosis research, in plants especially, is understanding the underlying mechanisms behind CO interference. There is considerable interest in being able to control this phenomenon and direct CO formation to specific regions of the genome, which are notoriously poor in CO events.

CO interference could rely on chromatin conformation and condensation levels, as well as genomic modifications including epigenetic marks. A considerable body of research on meiosis in plants has focused on understanding the different roles of proteins and underlying mechanisms involved in chromosome pairing, formation of the SC and resolution of the chiasmata into COs (Mercier *et al.* 2015). In contrast, little is known about the effect of epigenetics on plant meiosis, whether it concerns the expression of the genes involved or its involvement in chromatin structure and CO localisation on the chromosomes.

1.3. DNA-methylation and its role in plants

Genes and their various alleles are the blueprints of any biological mechanism. However, they are subjected to the activity of many additional processes which regulate when, where and which versions of these genes are expressed. Epigenetic markers encompass chemical changes to the structure and folding of DNA molecules and chromatin remodellers such as histones, which lead to heritable phenotypes between generations without altering the DNA sequence itself. These markers comprise (but are not limited to) DNA-methylation, histone modifications, including methylation and acetylation, histone variants and non-coding RNA (ncRNA) changes (Quadrana & Colot 2016). These epigenetic factors can lead to dynamic changes in chromatin structure and differential

accessibility of genetic material depending on the cell cycle stage, developmental stage, or in response to abiotic or biotic stresses. Epigenetic markers therefore constitute an additional level of regulation for gene expression. Major disruption of these markers usually leads to extensive developmental irregularities. Yet, there is a large interest in understanding how smaller, targeted changes could lead to differential gene expression benefitting genes of interest. Additionally, given the partially inheritable nature of these markers, there is growing awareness for elucidating the roles and mechanisms of epi-alleles, and how stably they are transferred between generations.

Most of the mechanics for DNA-methylation in plants have been studied using *Arabidopsis* as a model. For this reason, most of the genes discussed thereafter and their role will have been characterised in this model, as reviewed by Zhang *et al.* (2018). However, given that the DNA-methylation dynamic processes are very well maintained between species, most of the corresponding orthologues of these genes will work in an identical of very similar fashion in other plant species, even if named differently.

1.3.1. Mechanisms of DNA-methylation in plants

DNA-methylation mechanisms are largely conserved between organisms including plants and mammals. Each species is characterised by precise patterns of DNA-methylation that are generally inherited and indispensable for correct development (Law & Jacobsen 2010). In plants, DNA-methylation tends to be abundant in heterochromatic regions of the chromosomes, which are rich in transposable elements (TE) and highly repetitive DNA. However, clusters of DNA-methylation will also be found in euchromatic regions, especially in TE-rich regions sequences (Zhang *et al.* 2006, Eichten *et al.* 2013, Li *et al.* 2012, Takuno & Gaut 2013).

RNA-directed *de novo* methylation relies on the use of small interfering RNAs (siRNAs) to target the regions to be methylated. Maintenance of DNA-methylation is based on copying DNA-methylation patterns after DNA-replication using the original DNA strand as a model. DNA-methylation is always linked to cytosine sequence contexts CG, CHG and CHH, where H represents A, T or C (Takuno & Gaut 2013). Maintenance of DNA-methylation in each of these sequence contexts involves a different set of proteins for methylation, as well as for the demethylation processes. RNA-directed *de novo* methylation, however, uses the same pathways independently of the sequence context. Whichever DNA-methylation process is used, the methyl group will always be linked

either to position 5 of the pyrimidine ring of the targeted cytosine or to a lesser extent on an adenine in position 6 (Vanyushin & Ashapkin 2011). A methylated cytosine will therefore be referred to as a 5-methylcytosine, or 5-mC. DNA-methylation reactions are catalysed by various DNA-methyltransferases, which use S-adenosyl-L-methionine as a substrate and methyl-group donor. In contrast, DNA-demethylation reactions involve excision of the 5-mC base altogether using 5-mC glycosylases (Zhang *et al.* 2018).

1.3.1.1. De novo DNA-methylation

RNA-directed DNA-methylation (RdDM) is the biological process responsible for *de novo* DNA-methylation in plants and consists of several complex and intertwined pathways (Figure 1.3.1). In brief, siRNA transcription is initiated by RNA Polymerase IV (POL IV) in association with RNA-Dependent RNA Polymerase 2 (RDR2), which is involved in copying the transcript to produce a double-strand RNA (dsRNA). These dsRNAs are consequently cleaved into 24-nucleotide siRNAs by a range of Dicer-Like (DCL) proteins such as DCL2, DCL3 and DCL4 (Pikaard *et al.* 2012). The siRNAs are then transported to the DNA-methylation locations by Argonaute proteins (AGO) such as AGO4 and AGO6 (Gao *et al.* 2010). In parallel, scaffold RNAs are produced by RNA Polymerase V (POL V) transcription. These scaffold RNAs are then attached to the chromatin with the help of RNA-Processing-6-Like 1 (RRP6L1) (Zhang *et al.* 2014). A complex is formed between POL V, RNA-Directed DNA-methylation 3 (RDM3) and the siRNA-carrying AGO proteins which allows the binding of the siRNA with the scaffold RNA (Bies-Etheve *et al.* 2009). This complex finally interacts with Domains Rearranged Methylase 2 (DRM2), and this association in turn triggers the *de novo* DNA-methylation reaction. DNA-methyltransferase DRM2 is active independently of the DNA-sequence (Zhong *et al.* 2014).

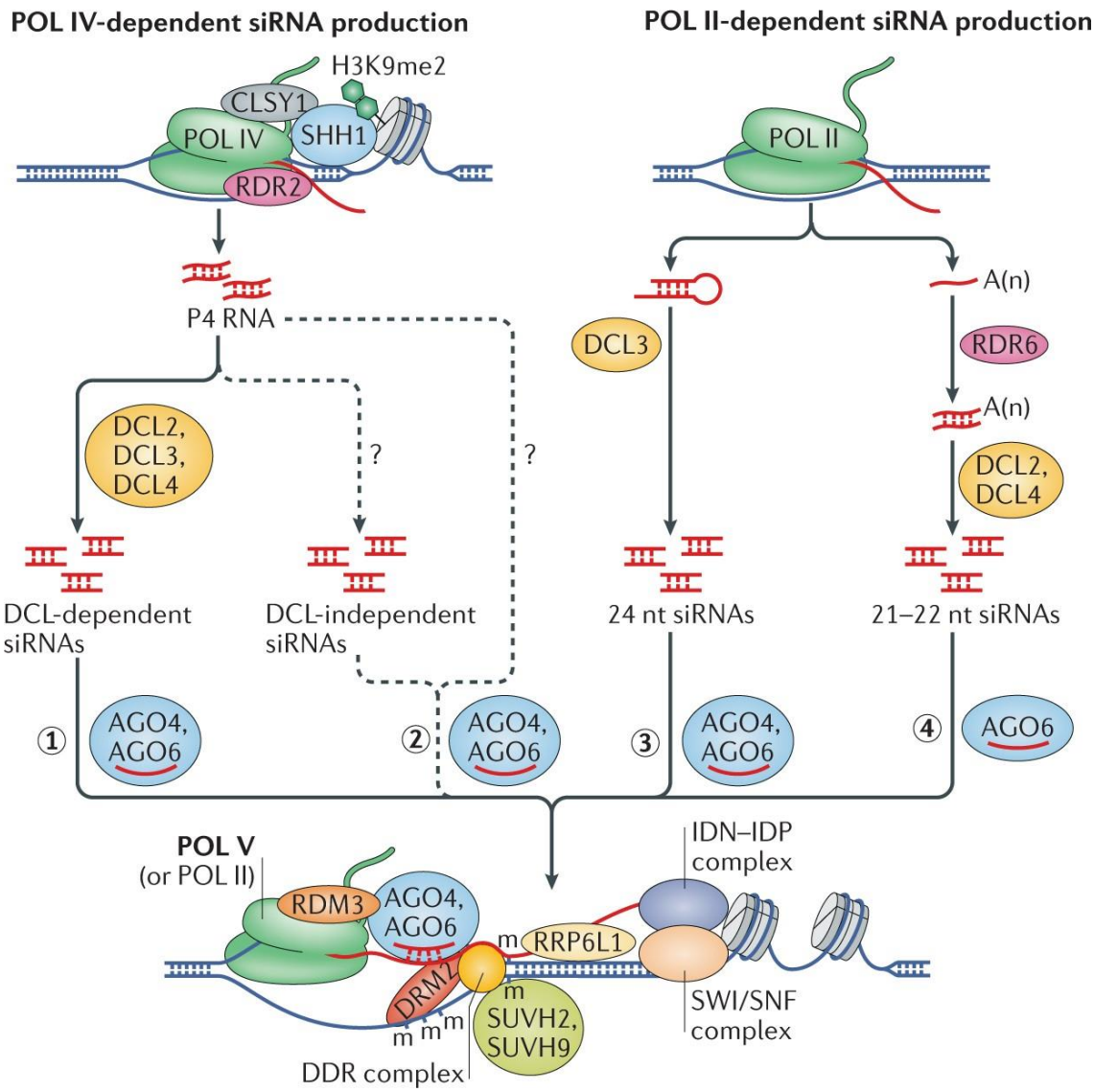


Figure 1.3.1. Simplified schematic of the RDdM pathway in plants shows production of siRNAs by POL IV These siRNA are cleaved by DCL proteins before being brought to the de novo methylation site by AGO proteins. Scaffold RNAs are transcribed by POLV and aligned to the target sequence by RRP6L1. This complex then recruits the DNA-methytransferase DRM2 which is active regardless of the DNA-sequence context. Adapted from Zhang et al. (2018).

Many other proteins are involved in the RdDM process, which help facilitate reactions or stabilise the complex. In particular, the DDR complex (DRD1, DMS3 and RDM1) is essential for the correct association of POL V with chromatin to produce scaffold RNA (Law *et al.* 2010, Kanno *et al.* 2004, Kanno *et al.* 2008, Gao *et al.* 2010). Moreover, the siRNA-scaffold RNA complex is stabilised by the binding of the IDN-IDP (Involved *De Novo* 2-IDN Paralogue), which in turns interacts with a SWI/SNF (Switch/Sucrose Nonfermenting) complex. SWI/SNF is involved in chromatin remodelling and will be involved in transcriptional silencing for POL V by altering the positioning of the nucleosome next to the polymerase (Ausin *et al.* 2009, Zhu *et al.* 2013).

Other secondary pathways for *de novo* methylation have also been described in *Arabidopsis*. These potentially involve cleavage of the dsRNA into siRNA by a DCL-independent pathway and are currently not fully understood. POL II can also be involved in producing siRNAs, which as for the POL VI-dependent siRNA production, will be cleaved by DCL proteins and carried to the DNA-methylation locus by AGO proteins. In addition to 24-nucleotide siRNAs, the POL II pathway seems to also produce 21 or 22-nucleotide siRNAs, though their difference in action is still unclear (Zheng *et al.* 2009).

1.3.1.2. Maintenance of DNA-methylation

The DNA-methyltransferases involved in replicative DNA-methylation of different cytosines are regulated by separate mechanisms depending on the sequence context (Figure 1.3.2).

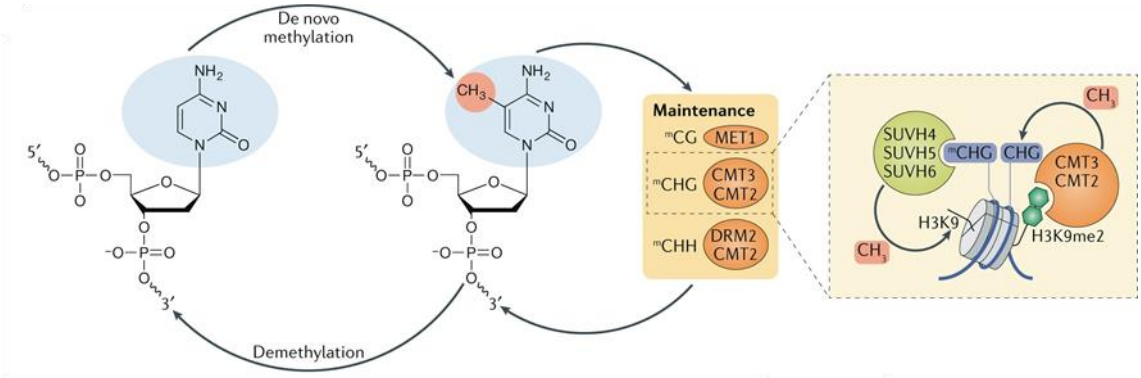


Figure 1.3.2. Overview of the maintenance of methylation process in plants. Maintenance of DNA-methylation occurs in all cytosine contexts but is catalysed by different methyltransferase depending on the sequence. CG-methylation is maintained by MET1, CMT2 and CMT3 are involved in CHG-methylation, whilst DRM2 and CMT2 direct maintenance of methylation in CHH context. Adapted from Zhang et al. (2018).

Methyltransferase 1 (MET1) recognises hemi-methylated CG nucleotides and is therefore involved in CG methylation (Kankel *et al.* 2003). This orthologue of the human DNA (cytosine-5)-methyltransferase 1 (DNMT1) will be able to methylate the demethylated cytosines in CG context in the daughter strand after replication. Variant In Methylation 1 (VIM1) and its orthologues have been proposed to be essential for the maintenance of CG methylation in *Arabidopsis* by recruiting MET1 to the chromatin. However, the mechanism is yet to be solved (Hye *et al.* 2007).

DNA-methylation of CHG sequences is catalysed mainly by Chromomethylase 3 (CMT3). Chromomethylase 2 (CMT2) can also be involved, but less frequently (Lindroth *et al.* 2001). Work on the CMT3 orthologue Chromomethylase 1 (MET2A) in maize showed the protein carries a bromo-adjacent homology (BAH) domain and a chromodomain which

bind to methylated histone H3K9me2 (Du *et al.* 2012). This interaction is crucial for correct binding of the methyltransferase to the nucleosome and therefore for appropriate CHG methylation site.

Maintenance of CHH methylation is regulated either by DRM2 mentioned above in *1.3.1.1*, or by CMT2, depending on the target region. DRM2 is involved in CHH methylation through the RdDM pathway, and CMT2 replaces the DRM2 pathway in regions where DRM2 is inhibited, especially in heterochromatin linked to H1 histones (Zemach *et al.* 2013).

Many other genes are potentially involved in the maintenance of DNA-methylation, and the exact interactions between the factors involved in methylation of the genome are still unclear. For example, Decreased DNA-methylation 1 (DDM1) is a chromatin-remodelling protein which has been proven to have an important role in maintaining levels of DNA-methylation in all pathways, including RdDM. However, its precise mode of action is still unclear (Zemach *et al.* 2013).

1.3.1.3. DNA-demethylation

Unlike DNA-methylation which involves DNA-methyltransferases adding methyl groups directly onto cytosines already incorporated into the DNA, active DNA-demethylation in plants requires excision of the bases carrying the methyl groups by glycosylases through a base excision repair pathway (Gong *et al.* 2002). Four glycosylases have been studied and characterised in *Arabidopsis*.: Repressor Of Silencing 1 (ROS1), Demeter-Like protein 2 and 3 (DML2 and DML3), and transcriptional activator Demeter (DME) (Gong *et al.* 2002, Ortega-Galisteo *et al.* 2008, Morales-Ruiz *et al.* 2006). They can excise the 5-mC bases regardless of the sequence context.

Briefly, the demethylases first hydrolyse the bond between the cytosine base and the deoxyribose sugar. They then exhibit a lyase activity to cut the DNA backbone and fully excise the rest of the nucleotide. The consequent base repair pathway is not fully understood and the polymerase involved remains undiscovered, however, it has been suggested that DNA Ligase 1 (LIG1) could be involved in the process (Yan Li *et al.* 2015).

DNA-demethylases target specific regions of the genome, depending on chromatin status and the sequence context of the region. For example, DME and ROS1 appear to preferentially target TEs situated near genes and may be involved in the differential expression of these genes (Gehring *et al.* 2009, Hsieh *et al.* 2009, Tang *et al.* 2016).

Moreover, some regions targeted by ROS1 need the recruitment of an anti-silencing Increased DNA-methylation (IDM) complex to facilitate the activity of ROS1. This complex ensures targeting of the DNA-demethylation process to hypermethylated regions, though its full way of functioning is yet to be determined (Qian *et al.* 2012).

In addition to active DNA-methylation processes, passive DNA-demethylation can also occur in cases where there is a lack of S-adenosyl-L-methionine as a methyl donor, or if the DNA-methyltransferase activity itself has been reduced. This also results in a failure to maintain DNA-methylation along the genome and will likely lead to hypomethylated regions or full genomes (Rocha *et al.* 2005, Zhang *et al.* 2012, Groth *et al.* 2016).

1.3.1.4. Methylstat mechanism

DNA-methylation is known to be involved in many biological processes and patterns of methylation on the genome are heavily maintained between individuals of a species and between generations. This suggests the balance between DNA-methylation and demethylation is tightly regulated. Extensive research has used mutants in genes involved in DNA-methylation, generated in *Arabidopsis*, in order to explain the mechanisms of this regulation.

As expected, *ros1* mutants exhibit DNA hypermethylation in many genomic regions. Surprisingly, *nrpd1* mutants are characterised by a global DNA hypermethylation, but most interestingly by a drop in ROS1 gene expression. These mutant lines carry a defective subunit of POL IV, the DNA-Directed Pol IV Subunit 1 (NRPD1), which is involved in the production of siRNAs. This suggests a mode of signalling between RdDM pathways and demethylation by ROS1. The low levels of ROS1 expression in these mutants could therefore explain the hypermethylation of the genome (Tang *et al.* 2016).

Inhibition of ROS1 expression could also be observed in *met1* mutants (Mathieu *et al.* 2007). It has been shown that the promotor of ROS1 is characterised by a 39 bp sequence, called the DNA-methylation Monitoring Sequence (MEMS). The MEMS is suggested to be an indicator of DNA-methylation levels in the genome (Lei *et al.* 2015, Williams *et al.* 2015). This region was discovered to be fully demethylated in *met1* mutants, as well as in other mutants involved in the RdDM pathway, which triggers a lower ROS1 expression. Furthermore, DNA hypermethylation can be observed in the MEMS in *ros1* mutants, with increased gene expression of the mutated ROS1, further implying that this promotor sequence is involved in sensing the activity of DNA-methyltransferases and adapting

ROS1 expression accordingly. MEMS could therefore be considered to act as a methylstat, sensing and maintain stable levels of DNA-methylation in the genome. This methylstat balances ROS1 mechanisms for demethylation, making it responsible for the homeostasis of methylation levels in the genome. Interestingly, loss of CG methylation in *met1* mutants is compensated by an increase in CHH methylation between generations, resulting in ROS1 expression being silenced again with time (Mathieu *et al.* 2007).

Similar processes are likely to exist for the other proteins involved in DNA-methylation, but their existence and mode of functioning has not yet been clarified. Such mechanisms have also been observed in other species such as rice and maize (Williams *et al.* 2015, Hu *et al.* 2014, Erhard *et al.* 2015), suggesting this methylstat might be conserved among plant species and crucial in regulating the balance between DNA-methyltransferases and demethylases.

1.3.2. Roles of DNA-methylation in plant development

As one of the major epigenetic markers, DNA-methylation, is strongly involved in developmental mechanisms such as gene expression and TE silencing by modifying chromatin structure and accessibility.

1.3.2.1. Gene expression regulation

In most cases, DNA-methylation in gene promotors is involved in the repression of transcription of the gene, either by inhibiting the binding of transcriptional activators or promoting the binding of transcriptional repressors. There are also cases where DNA-methylation is involved in promoting the transcription of a gene. For example, DNA-methylation is involved in activating genes involved in repressing fruit-ripening in tomato, preventing the fruits from maturing too quickly post-harvest (Lang *et al.* 2017). However, the mechanisms underpinning DNA-methylation activated gene expression are still unclear. In these cases, DNA-methylation could be involved in recruiting transcription factors or preventing the binding of repressors of transcription, but the exact mechanisms are still to be elucidated. DNA-methylation mediated gene expression is also regulated by the activity of DNA-demethyltransferases which will either prevent transcriptional silencing in the case of genes repressed by methylated promotors, or repress transcription if DNA-methylation is necessary for correct gene expression (Liu *et al.* 2015).

Gene bodies can also be methylated, mostly within CG sequence contexts. DNA-methylation targets mostly exons and is absent from transcription initiation sites and stop codons (Cokus *et al.* 2008). Patterns of gene body methylation (gbM) seem to be highly conserved between plant species, preferentially targeting longer genes, with more exons, which are constitutively expressed (Takuno & Gaut 2013).

Despite gbM patterns being highly conserved between plant species, the role of methylation within genes seems to be highly species dependent. In contrast to gene promotors, over a third of the genes in *Arabidopsis* are methylated (Zhang *et al.* 2006). These genes are mostly demethylated in demethylating mutants, however, transcript levels for these genes do not seem to be impacted. This suggests that in *Arabidopsis* gbM does not affect expression levels of the genes. This is further confirmed by the fact that gbM patterns variation between different accessions of *Arabidopsis* is not correlated with differential gene expression (Kawakatsu *et al.* 2016). In contrast, in rice (*Oryza sativa*) *met1-2* mutants that are characterised by a global loss of CG-methylation, including in gene bodies, a portion of the genes appear to be alternatively spliced (Wang *et al.* 2016). This suggests gbM could improve pre-mRNA splicing efficiency in some plant species.

1.3.2.2. Transposon regulation

As described in 1.3.1, DNA-methylation islands are mainly located in TE rich regions. This suggests DNA-methylation might have a role in the repression of TEs. This is of crucial importance for plant development and maintenance through generations as this ensures genome stability is maintained by repressing the risk of relocation of DNA transposons or the insertion of retrotransposons in active genes or promotors.

It is unclear how DNA-methylation ensures silencing of TE. In maize, loss of CHH methylation islands leads to higher transcription levels for the affected genes. Neighbouring TEs are also characterised by demethylated patterns (Qing Li *et al.* 2015). In rice demethylase activity promotes TE activity and re-localisation. Here, a knock-out mutation of the ROS1 orthologue DNA-Glycosylase/Lyase 701 (DNG701) led to a reduction in the activity of the retrotransposon Tos17 (La *et al.* 2011). In contrast, in *Arabidopsis*, even though demethylated mutants showed de-repression of TE, very few transposition events were observed. In double mutants for both *met1* and *cmt3* (Kato *et al.* 2003), in single *met1* mutants (Mirouze *et al.* 2009), or in *ddm1* mutants (Tsukahara *et al.* 2009), levels of transposition were slightly elevated. Some transposition events

were observed in RdDM nrpd1 mutants under heat stress conditions, but not in normal growth conditions (Ito *et al.* 2011). Collectively these observations indicate that CHH methylation, even though it is the main methylation context in TE, is not involved in transposition silencing on its own.

1.3.2.3. DNA-methylation and chromatin interactions

By modifying DNA structure and interacting with histones, DNA-methylation is involved in chromatin structure and therefore, chromosome interactions. In *Arabidopsis*, the KNOT structure, named after the unsolvable Gordian Knot, describes the particular interactions of all five chromosomes in somatic nuclei (Grob *et al.* 2014). The chromosomes interact via interactive heterochromatin islands (IHIs) (Feng *et al.* 2014). These IHIs are located in euchromatin regions but are enriched in TEs and small RNAs. Unexpectedly, *met1* and *ddm1* mutant plants in *Arabidopsis* were characterised with ectopic IHI loci, and no diminution in chromosome interactions, despite extensive demethylation in all cytosine contexts. Additionally, RdDM mutants showed increased frequency in IHIs, which again would suggest RdDM would be involved in preventing chromosome interactions via IHI in specific regions of the genome (Rowley *et al.* 2017). In contrast, *met1* and *ddm1* mutants in *Arabidopsis* see a reduction in pericentromeric interactions not involved in the KNOT complex (Feng *et al.* 2014). This results in a re-localisation of interactive events within peri-centromeric regions. Moreover, *ddm1* and *met1* mutants are characterised by a decrease in interactions between pericentromeric regions themselves, but an increase in interactions between pericentromeric and euchromatic regions of chromosome arms (Feng *et al.* 2014).

Research in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and *Caenorhabditis elegans* indicates that DNA-methylation plays a role in the final location of CO events, as reviewed by de Massy (2013). In many crops such as tomato, wheat or barley, the large highly methylated heterochromatic regions are also almost completely devoid of COs (Sato *et al.* 2012, Choulet *et al.* 2014, Mayer *et al.* 2012). Most CO events occur in distal euchromatic regions, which tend to be globally demethylated. This suggests that a chromosomes' primary structure likely plays a role in the final CO landscape in plants.

In *Arabidopsis*, *ddm1* and *met1* mutants exhibit shifts in their meiotic recombination landscapes. In particular, in *ddm1* mutants, the observed meiotic recombination rate is higher in euchromatic regions compared to wild type individuals, with no difference

between male and female meiocytes, but no effect was observed on the number of CO events in pericentric regions (Melamed-Bessudo & Levy 2012). Similarly, a study of *met1-3* mutants in *Arabidopsis* by Yelina *et al.* (2012) observed a decrease in CO rates in pericentromeric regions, whereas higher CO rates were noted in both centromere-proximal and distal regions. In addition to the role of DNA-methylation maintenance genes on meiosis, RdDM genes are also involved in the meiotic process by regulating gene expression in meiocytes (Walker *et al.* 2018). More relevant to large genome crops, Virus-Induced Gene Silencing (VIGS) of *met1* and *ddm1* in wheat led to increased levels of CO events in the sub-telomeric region of the short arm of chromosome 1A, suggesting the effects seen in *Arabidopsis* can also been observed in cereals (Raz *et al.* 2021). The specific roles of these genes involved in DNA-methylation and their effect on meiotic recombination will be further discussed in Chapter 4.

1.4. Project aims.

As explained in 1.1.3, the rate of genetic gain in plant breeding is currently stalling. Genetic modification is still largely unaccepted by members of the general public, and it is currently illegal to grow genetically modified plants for commercial use in Europe. Traditional breeding is limited by recombination rates, which in large genome crops are highly skewed towards the sub-telomeric ends of the chromosomes, excluding large portions of the genome from the breeding process. Therefore, there is considerable interest in understanding how recombination events are regulated and in controlling where COs happen in the genome. Consequently, meiosis in plants has been extensively studied in order to resolve the underlying mechanisms which control chromosome pairing and recombination events. However, very little has been investigated about the effect of epigenetic markers on meiotic recombination. My PhD project aimed to explore the impact of DNA-methylation on CO location in barley and *Arabidopsis*. This subject was focused around three major areas of research:

- Can tools be optimised to fit the needs of studying the role of DNA-methylation in *Arabidopsis* and barley?

  Hypotheses: *A pre-existing protocol for the assessment of DNA-methylation in human liver cells can be optimised to identify hypomethylation in plant material. Pre-existing RNAseq data on gene expression levels in* Arabidopsis *plantlets treated with 5-azacytidine, a zebularine analogue, can help identify genes which*

are differentially regulated in demethylated plants; these genes can be used as markers in Arabidopsis *zebularine-treated seedlings.*

- Does the application of zebularine during germination impact meiotic recombination in *Arabidopsis* and barley?

  Hypothesis: *Similar to what had been observed previously in met1-3 and ddm1 mutants in* Arabidopsis*, zebularine-treated seedlings of* Arabidopsis *and barley will exhibit altered CO frequency in sub-telomeric and peri-centromeric regions of the chromosomes.*

- Can mutations in methylation genes *met1* and *ddm1* help redefine the meiotic landscape in barley (*cv.* Golden Promise)?

  Hypothesis: *Reverse and forward genetics approaches will lead to the identification of a range of mutations in* met1 *and* ddm1 *which will cause various hypomorphic effects on plant development and meiotic recombination.*

2. General Materials and Methods

2.1. Plant Material

2.1.1. *Arabidopsis thaliana*

Arabidopsis were grown at the Plant Sciences department at the University of Cambridge in 9 cm square pots with a density of 6-9 plants per pots depending on application and size. Plants were transferred to controlled growth chambers (20°C, 60% humidity, 16 hours day, 8 hours night) after a stratification period of 4 days at 4°C. Plants were grown in a standard commercial soil (F2) supplemented with vermiculite. These growth rooms were suitable for the cultivation of Genetically Modified (GM) plant material and were used following the official Health and Safety regulations relative to the handling of GM material.

2.1.2. Barley

Barley (*cv.* Golden Promise, Barke, and Jettoo F1 Hybrid) was grown in a controlled greenhouse environment (16 hours day, 8 hours night, 18-15°C) in a range of pot sizes and multiple-cell trays depending on their intended use. Generally, larger pots were used for crossing plants and when a larger seed yield was required, and trays were used when plants were mainly used for DNA extractions and genotyping. Seeds were germinated directly in the pots in ¼ strength cereal-specific mix with added insecticides (To make around 8kg: 2.5 kg Dolomitic limestone, 2.5 kg Garden limestone, 1.5 kg Osmocote® Start 6 weeks, 875 g Osmocote® Exact Standard 3-4 months, 500 g Celcote®, 390 g Intercept®) and the plants were placed onto automatic watering benches.

2.2. Molecular Biology

2.2.1. Barley and *Arabidopsis* DNA extractions

When processing small numbers of samples at the same time, DNA extractions were performed using the DNeasy Mini Kit (QIAGEN) following the supplier's protocol.

Briefly, plant material (less than 100 mg) was recovered in 2 mL DNase-free Eppendorf tubes either from leaves or full seedlings and flash frozen in liquid nitrogen in the presence of five to seven 3 mm tungsten beads. The material was then disrupted by vortexing the flash-frozen tubes together with the metal beads until a fine powder was obtained. The DNA extraction process was then carried out as described by the

manufacturers' protocol. However, the final elution step was modified as follows: 100 µL of AE buffer was added to the centre of the column and the samples were left to incubate for 5 minutes at room temperature before being centrifuged for 1 minute at 6,000 x g. To maximise yields, the eluted solution was pipetted back into the column, left to stand for another minute and centrifuged again for 1 minute at 6,000 x g.

When extracting DNA from large numbers of individuals, leaf material (around 50 mg) was supplied in plates to the genomics facility at the James Hutton Institute where DNA extraction was performed as a routine service.

2.2.2. Polymerase Chain Reactions (PCR)

PCR was used during this project to amplify DNA fragments of interest. When the PCR was carried out with the aim of using the amplicons as templates for sequencing (Sanger or MiSeq), proof-reading DNA polymerases Q5 or KAPA were used. When a diagnostic PCR was carried out to confirm the presence/absence of a DNA sequence or to new primers, GoTaq® G2 Flexi DNA Polymerase (Promega) was used.

2.2.2.1. Q5 Hi-Fi Hot Start DNA Polymerase (New England Biolabs)

PCR reactions were as follows: 5 µL Q5 Reaction Buffer (5X), 0.5 µL dNTPs (10 mM), 0.25 µL Q5 Hi-Fi Hot Start DNA Polymerase (2U/µL), 1.25 µL Forward/Reverse primers (10 mM), 2 µL[1] DNA template and sterile water to complete to a volume of 25 µL per reaction.

The samples were then submitted to the following PCR program on an AC1 Alpha thermal cycler (PCRmax): Initial denaturation at 98°C for 30 seconds, 30 cycles of denaturation at 98°C for 10 seconds, annealing at 62°C for 15 seconds and extension at 72°C for 30 seconds, final extension at 72°C for 2 minutes, holding stage at 8°C.

Underlined volumes, temperatures and times were adjusted to the needs of the specific experiments.

2.2.2.2. KAPA HiFi HotStart ReadyMix PCR Kit (Roche)

PCR reactions using the KAPA ReadyMix PCR kit were prepared to a final volume of 25 µL for each reaction according to the following quantities: 12.5 µL KAPA HiFi HotStart

[1] Variable quantities, temperatures and extension times in PCR reactions are underlined.

ReadyMix (2X), 0.75 µL Forward/Reverse primers (10 mM), 2 µL DNA template and sterile water to make up to the final volume of 25 µL.

The PCR program was carried on the thermal cycler according to the following sequence: Initial denaturation at 95°C for 3 minutes, 30 cycles of denaturation at 98°C for 20 seconds, annealing at 62°C for 15 seconds and extension at 72°C for 30 seconds, final extension at 72°C for 3 minutes, holding stage at 8°C.

Underlined volumes, temperatures and times were adjusted to the needs of the specific experiments.

2.2.2.3. GoTaq® G2 Flexi DNA Polymerase (Promega)

Reactions using GoTaq® G2 Flexi were made as follows: 5 µL Green or Colorless GoTaq® Flexi Buffer (5X), depending on the application, 2 µL MgCl2 (25 mM), 0.5 µL dNTPs (10 mM), 0.25 µL GoTaq® G2 DNA Polymerase (5 U/µL), 1.25 µL Forward/Reverse primers (10 mM), 2 µL DNA template and sterile water up to a final volume of 25 µL.

Templates were amplified using the following PCR program: Initial denaturation at 94°C for 2 minutes, 30 cycles of denaturation at 94°C for 15 seconds, annealing at 62°C for 20 seconds and extension at 72°C for 30 seconds, final extension at 72°C for 2 minutes, holding stage at 8°C.

Underlined volumes, temperatures and times were adjusted to the needs of the specific experiments.

2.2.3. Column-based PCR Purification

PCR product purification was achieved using the QIAquick PCR Purification Kit (QIAGEN) following the protocol provided by the manufacturer. Briefly, 5 volumes (125 µL) of PB buffer were added to the PCR products post-PCR and the solution was mixed by pipetting up and down. The mixture was then transferred to a fresh QIAquick column on a collection tube before being centrifuged for 1 minute at 14,500 x g. The flow-through was discarded and the samples were washed on the column with 750 µL PE buffer. After 1 minute centrifugation at 14,500 x g, the flow-through was discarded and the empty column was again centrifuged for 1 minute at 14,500 x g to dry the membrane. The column was then placed on a fresh 1.5 mL collection tube and 50 µL of sterile water were added in the centre for the elution step. The column was left to stand for 1 minute before being centrifuged at 14,500 x g for 1 minute. On some occasions where a higher yield was

needed, the eluted DNA was placed back into the column, left to stand for a further minute at room temperature and the column was finally centrifuged again for 1 minute at 14,500 x g.

2.2.4. Barley and *Arabidopsis* RNA extractions

Depending on the starting material, the desired quality, and the application for which the RNA samples were used, two methods of RNA extractions were used. The RNeasy kit (QIAGEN) was adapted to extract RNA from barley seedlings and traditional phenol/chloroform RNA extractions were performed using TRIzol reagent (Invitrogen) on *Arabidopsis* seedlings, leaf tissue and barley leaf tissue.

2.2.4.1. RNeasy Plant Mini kit (QIAGEN)

The supplied protocol from the manufacturer was adapted by adding an on-column DNase treatment, and to suit the needs required by *Arabidopsis* material. This protocol was developed by a previous PhD student (Mittmann, 2017) and again optimised for the plant material used.

A maximum of 100 mg of plant material was collected in 2 mL Eppendorf tubes with steel beads and was flash-frozen in liquid nitrogen. Whilst still frozen, the tubes were vortexed at high speed until a fine powder was obtained. Immediately after disruption, 450 µL of freshly prepared Buffer RLT (10 µL β-mercaptoethanol for 1 mL Buffer RLT) was added to the powder and vortexed vigorously before transferring the mix to a fresh QIAshredder column. Samples were centrifuged for 2 minutes at maximum speed. The supernatant was then transferred to a fresh RNase-free 1.5 mL tube without disturbing the pellet and 225 µL of fresh ethanol (96-100%) was added and mixed by pipetting. The sample was then immediately transferred to a fresh RNeasy spin column and centrifuged for 30 seconds at 8,000 x g. After discarding the flow-through, 350 µL Buffer RW1 was added to the centre of the column. The sample was centrifuged for 30 seconds at 8,000 x g and the flow-through was discarded. DNA degradation was carried out on the column by adding a mix of 70 µL of RDD Buffer and 10 µL of DNase I (QIAGEN) to each sample and leaving it to incubate at room temperature for 20 minutes. 350 µL Buffer RW1 was then added, and the columns were centrifuged for 30 seconds at 8,000 x g. After discarding the flow-through, 500 µL Buffer RPE was added to the column and the samples were centrifuged at 8,000 x g for 30 seconds. The latter step was repeated once for a second wash and the empty columns were then centrifuged at 16,000 x g for 2 minutes to

remove the excess ethanol. The column was then transferred onto a fresh 1.5 mL RNase-free tube and 50 µL of RNase-free sterile water was added in the centre of the column for elution and the samples were centrifuged at 8,000 x g for 1 minute. To maximise RNA concentrations, the eluted RNA was pipetted back onto the column and re-centrifuged for 1 minute at 8,000 x g.

2.2.4.2. Trizol based RNA extraction.

Trizol RNA extraction was used to obtain high quality RNA in large enough quantities for microarray analysis of barley gene expression. RNA isolation was carried out following the manufacturers' instruction.

Briefly, around 100 mg of barley leaf material was flash-frozen in liquid nitrogen in 2 mL Eppendorf tubes with stainless steel beads and were immediately vortexed at high-speed whilst frozen until a fine powder was obtained. The samples were resuspended in 1 mL of Trizol and immediately homogenised by vortexing. After a 5-minute incubation at room temperature, 200 µL of cold chloroform was added to the samples. The tubes were then vigorously shaken, and the samples were incubated a further 3 minutes at room temperature. The tubes were then centrifuged for 15 minutes at 4°C at 12,000 x g, leading to a separation of the sample mixture into three phases: a red phenol-chloroform phase at the bottom of the tubes, a viscous interphase, and a clear upper phase. The upper aqueous phase was transferred to a fresh RNase-free 1.5 mL Eppendorf, and RNA precipitation was performed by adding 500 µL of cold isopropanol to the samples, incubating the tubes for 10 minutes on ice, and centrifugating the samples for 10 minutes at 4°C at 12,000 x g. The supernatant was then discarded, and the RNA pellet was washed with 1 mL of freshly prepared 75% ethanol. The pellet was resuspended by briefly vortexing the tubes and was then re-precipitated by a 5-minute centrifugation step at 4°C, at 7,500 x g. The supernatant was thoroughly removed, and the RNA pellet was allowed to dry at room temperature for 10 minutes until all ethanol had evaporated. The RNA pellet was then resuspended in 20 µL RNase-free water and incubated for 10 minutes at 55°C to allow it to dissolve completely. Given the nature of the microarray assay which does not rely on sequencing and cDNA amplification, no DNase treatment was performed on the samples.

2.2.5. DNA/RNA Gel Electrophoresis

Extracted DNA and RNA as well as PCR products or digestion products were analysed by agarose gel electrophoresis on 0.7 – 2 % agarose gels including ethidium bromide (0.5 µg/mL final concentration). Depending on the application, 5 – 20 µL of each sample was mixed with an equal volume of 2X Blue/Orange loading dye (Promega) before being loaded on the gel. Gels were run at 80 – 120 V for 30 – 60 minutes depending on the agarose concentration and the type of product being analysed. 10 µL of 100 bp or 1 kbp ladder (Promega) was used for size comparison. Gels were then imaged using UV transillumination. In the case of RNA products, all electrophoresis material and all glassware were previously cleaned and treated against RNase using disinfectant and RNaseZap™ (Invitrogen).

2.3. Bioinformatics

2.3.1. Primer design

Primers destined to be use for PCR amplifications were designed using genomic DNA as a template (*cv.* Golden Promise) using Primer3 Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi, Untergasser et al., 2007). To design RT-qPCR primers, the Universal ProbeLibrary Assay Design Center (Roche, https://lifescience.roche.com/en_gb/brands/universal-probe-library.html#assay-design-center) was used with cDNA, selecting *Arabidopsis* as the target organism. Finally, WASP (http://bioinfo.biotec.or.th/WASP, Wangkumhang et al., 2007) was used as a tool to design allele specific primers and SNP-based primers on barley genomic DNA.

2.3.2. Sanger sequencing

DNA sequencing was carried out as a service by the genomics facilities at the James Hutton Institute. In all cases, in house primers were used and provided to the genomics facility. Sequencing results and quality were analysed using Chromas 2.6.5 (Technelysium) and alignments were done either using Multalin (Corpet 1988) or Mega 7.0.26 (Kumar *et al.* 2016).

3. Developing tools to study DNA-methylation in *Arabidopsis thaliana* and barley

3.1. Introduction

This chapter describes the development of a range of tools and methods used in the rest of this thesis to study the role of DNA-methylation on plant development and meiotic recombination. This introduction will focus in a first part on methods to measure DNA-methylation levels and in a second part on the role of RNAseq in gene expression analysis.

3.1.1. DNA-methylation determination methods

As described in Chapter 1.4, DNA-methylation of the genome is characterised by the addition of a methyl group on the 5th carbon of the pyrimidine ring of cytosine nucleotides. In plants, DNA-methylation occurs in 3 different contexts: CG, CHG and CHH where H represents A, C or T (Zhang *et al.* 2018). The choice of a practical method to measure levels of DNA methylation in a living organism depends on the expected outputs (*de novo* sequencing or detection of changes), on the target (whole genome or gene-specific), on the quality and amount of DNA template available, on the required sensitivity and robustness of the assay and, in the case of plants, on the target DNA-methylation context (global methylation or CG/CHG/CHH-specific) (Kurdyukov & Bullock 2016). Naturally, additional factors will influence the choice of method to be used, such as the availability of specific equipment, the simplicity of the method, the ease of handling the produced data and the overall cost of the assay itself. This section will describe the advantages and limitations of the main DNA-methylation assays currently available.

3.1.1.1. Bisulfite conversion-based methods

Bisulfite sequencing and the associated DNA-methylation assays based on bisulfite conversion of genomic DNA is considered by many as the standard method of measuring DNA-methylation levels across all living organisms. The method is based on the conversion of cytosines to uracil by deamination through a bisulfite-catalysed chemical reaction (Zhang *et al.* 2009). However, this reaction is methylation-sensitive and will leave

methylated cytosines intact. After PCR, un-methylated cytosine nucleotides will be read as thymidine, whereas methylated cytosine will continue to be read as cytosine. PCR products are then sequenced and after alignment, variant calling will allow the identification of thymidine nucleotides, which were annotated as cytosine prior to the bisulfite conversion. Data analysis is therefore simplified to the identification of Single Nucleotide Polymorphisms (SNPs) in the sequencing data, which result from the initial bisulfite conversion.

Similar to traditional Whole Genome Shotgun Sequencing, Whole Genome Bisulfite Sequencing (WGBS) represents the most comprehensive way of assessing DNA-methylation landscapes. This method allows for the assessment of the levels of DNA-methylation in the whole genome in general, but also in a context-specific manner. This is the method of choice when first characterising the methylome of a species as it allows for the identification of differentially methylated regions of genomes such as gene bodies, promoters, or transposon elements. WGBS can however present limitations depending on the species it is used for, both in terms of complexity of Next Generation Sequencing (NGS) data analysis and costs. A genome with a reduced proportion of cytosines, as is the case after bisulfite conversion, is less complex than the regular genome. This means the handling and alignment of sequencing data is complex, especially for larger genomes which already contain many repetitive sequences, such as large genome crops and many other plants (Li *et al.* 2013). In parallel, larger genomes necessitate more complex sequencing strategies, which can lead to much higher sequencing costs, making WGBS cost-prohibitive for many species.

An alternative was developed in 2005 (Meissner *et al.* 2005) allowing for analysis of 1% of a species' methylome, making data analysis and cost of the assays more accessible. The Reduced Representation Bisulfite Sequencing (RRBS) assay is based on the digestion of genomic DNA by a methylation insensitive enzyme, traditionally *MspI*, which recognises palindromic restriction sites 5'-C^CGG-3'. This leads to each DNA fragment carrying cytosine overhangs on each side, which can be methylated or not. The fragments are then submitted to an end-repair and A-tailing process, allowing for the ligation of sequencing adapters on each side of the fragments. Fragment purification allows selection of fragments of the desired size, which will vary depending on the species and the application of the RRBS method. Similar to WGBS, the fragments are then treated

with a bisulfite conversion chemical reaction. This leads to the extreme cytosines on each side of the fragments being deaminated into uracils, provided these cytosines were demethylated. Methylated cytosines will remain unchanged. The fragments are then amplified using primers specific to the adapters and sequenced using Sanger Sequencing or NGS depending on the selected size of the fragments. Due to the initial *MspI* digestion of the genomic DNA, all fragments will start with a C if the original CpG dinucleotide was methylated, or a T as a result of the conversion of an unmethylated Cytosine to an uracil during bisulfite conversion. This method leads to the enrichment of CG sequence context, leading to a decrease in the need for sequencing and a reduction in the overall cost. However, this method relies on the relative abundance of *MspI* restriction sites in the target genome, or of any similar enzyme resulting in a CG overhang on each side of the fragments. This means some DNA-methylation sites will not be identified, especially in large genome plant species, which contain many CHH and CHG methylation sites. Therefore, RRBS is better suited to the analysis of DNA-methylation for well-described, targeted regions of the genome of interest.

Finally, for the determination of the methylation status of specific genes or small genomic regions, bisulfite conversion can be associated to Methylation-Specific PCR (MSP) (Herman *et al.* 1996). After bisulfite conversion of the target genome, a PCR reaction is carried out using primers which amplify the region of interest and are specific to the methylation status of the target region. Primers are designed either to be specific to unconverted sequences which include cytosines, in which case they will not amplify if the target region is demethylated, or on the contrary to be specific to the converted cytosines to thymines, leading the PCR reaction to be unmethylated-specific. This method is particularly suited to regions of the genome which are highly methylated, as the increase in potential conversion events will lead to an increase in primer specificity. This methylation assay can also be made quantitative using quantitative PCR (MethyLight) (Trinh *et al.* 2001) or melting curve analysis (Mc-MSP) (Akey *et al.* 2002).

3.1.1.2. HPLC-based techniques

The high-performance liquid chromatography-ultraviolet (HPLC-UV) technique for DNA-methylation analysis relies on the distinction between deoxycytidines (dC) and methylated cytosines (5mC) (Beranek *et al.* 1980). Genomic DNA is first hydrolysed to separate the different nucleotide bases and chromatography used to quantify the

5mC/dC ratio in the DNA samples. This relative quantification method does not require a high level of expertise and allows for routine comparison between samples in various conditions. However, it requires a large amount of initial gDNA, does not allow for sequence specific methylation analysis and is limited by the need for specific equipment. This technique also does not allow for the detection of subtle changes in global DNA-methylation levels, limiting it to the detection of large effects on DNA-methylation only.

Sensitivity to the HPLC-UV assay can be increased by coupling liquid chromatography with mass-spectrometry (Song *et al.* 2005). Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) allows for the detection of more subtle changes in DNA-methylation in mammalian genomes. This technique presents an advantage over the traditional HPLC method in that it requires less DNA to proceed to the assay and is not impaired by lower-quality DNA. However, similar to the HPLC-UV assay for DNA-methylation, this method is limited by the equipment needed, as well as the required expertise to generate and analyse the data.

3.1.1.3. Immunoprecipitation-based methods

Methylated DNA immunoprecipitation (MeDIP) was first described in 2005 (Weber *et al.* 2005) when it was used to profile DNA-methylation levels in human chromosomes. This led to the observation that in female cells, the inactive X chromosome is globally hypermethylated compared to the other X chromosome. This technique relies on the use of antibodies raised against 5mC such as Methyl CpG Binding proteins MeCP2 and MBD2. The precipitated DNA can then be analysed and quantified either by DNA-microarray (MeDIP-chip) or by NGS (MeDIP-seq) (Kurdyukov & Bullock 2016).

During a MeDIP-chip assay, both the initial non-enriched DNA and immunoprecipitated DNA are labelled with distinct fluorescent probes and hybridised to a genomic microarray. The fluorescence ratios for the targets present on the microarray allow distinction between hybridisation levels of methylation-enriched and initial genomic DNA, allowing identification of hypermethylated regions in the samples. This method is relatively quantitative and is not particularly complex. However, it restricts the analysed targets to that of the ones already present on the probes and is therefore not suited to *de novo* methylation profiling.

In contrast, immunoprecipitated hypermethylated DNA can be purified, the antibodies removed, and the resulting DNA used for short-read sequencing. This MeDIP-seq method

allows for detection of novel hypermethylated islands, similarly to WGBS, but does not provide the same SNP resolution. Moreover, this technique is heavily biased towards hypermethylated regions of the genome, providing little to no information on genomic regions where DNA-methylation is present at much lower levels.

3.1.1.4. Enzyme based assays.

Most enzyme-based assays for quantifying DNA-methylation levels are based on pairs of restriction enzymes, which recognize the exact same restriction sites in the genome but are differentially sensitive to DNA-methylation. A few examples of such restriction enzymes are pairs *ClaI*/*BspDI* (5'-AT^CGAT-3') or *BspT104I*/*SfuI* (5'-TT^CGAA-3'), with the most commonly used enzymes being *MspI* (not methylation sensitive) and *HpaII* (methylation sensitive) which both recognise and cut restriction site 5'-C^CGG-3'.

The *HpaII* tiny fragment Enrichment by Ligation-mediated PCR (HELP) assay relies on the differential digestion of genomic DNA by *MspI* and *HpaII* in order to create fragments of sizes ranging from 200 bp to 2,000 bp (Khulan *et al.* 2006). These fragments can then be analysed either by short-read sequencing, in the case of a particular region of the genome being studied or hybridised to a microarray carrying the target sequences of interest. DNA-methylation patterns are then analysed by comparing fragment sizes between *MspI*- and *HpaII*-digested samples. Methylated restriction sites will result in longer fragments when digested with *HpaII* compared to *MspI*-digested samples.

Similarly, Luminometric Methylation Assays (LUMA) are based on *MspI*/*HpaII* restriction digestion coupled with pyrosequencing. Again, gDNA is separately digested by *MspI* and *HpaII* and the digested samples are submitted to the pyrosequencing process which will highlight any exposed C/G overhang. The ratio of *HpaII*-/*MspI*-related pyrosequencing readings then provides a measure of the DNA-methylation levels in the samples. As this assay does not require any modification of the gDNA such as bisulfite conversion, results can be obtained rapidly, and routinely. However, this technique requires the use of a pyrosequencer, limiting its use to a few laboratories only.

3.1.2. Gene expression analysis by RNAseq

3.1.2.1. Principles of RNAseq

RNA sequencing (RNAseq) relies on the sequencing and quantification of messenger RNA (mRNA) after extraction and reverse transcription. This can be done at various scales

depending on the size of the transcriptome, from whole organism tissues to even single cells.

The RNAseq process is initiated by the creation of an RNA sequencing library. In brief, total RNA is isolated from a tissue in the presence of DNase to remove any trace of genomic DNA. The total RNA is then submitted to a reverse transcription step leading to the production of cDNA. DNA is much easier to handle, store, and amplify compared to RNA, which makes the sequencing process possible. Sequencing libraries are then prepared in a very similar fashion to DNA sequencing libraries, often by fragmenting the cDNA molecules into smaller fragments before adding sequencing adapters on each extremity of the fragments.

After sequencing, reads are aligned and mapped. This can either be *de novo* if no reference genome is available for the species of interest or can be genome guided if a reference genome or reference transcriptome is readily available. Of course, *de novo* assembly is more challenging as it makes the process of building contigs and correcting sequencing errors more difficult. This also makes the identification of alternative splicing alleles of some genes more difficult as there is no full gene sequence available for the genes of interest.

Quantification of gene expression from the RNAseq data can be performed by comparing sequencing depth and read counts between genes and between samples from different conditions. Of course, such measures of gene expression are not absolute, as read counts will be normalised across samples to account for experimental variability when handling the samples and preparing the libraries. Normalisation is traditionally carried out using counts per million mapped reads for each analysed gene. Similarly, gene expression changes between samples in different conditions will be evaluated using fold change logarithm using the control sample as a reference. A positive log fold change represents a gene being over expressed in the sample, whereas a negative log fold change represents a downregulation of the gene. The log fold change is also a measure of the extent of the effect of the differential gene expression. Log fold change values are finally associated with a p-value as a measure of the statistical significance of the log fold change. Depending on the nature of the RNAseq study, different p-values can be used, which will dictate the threshold at which a gene will be considered as differentially expressed between samples.

3.1.2.2. 3D RNAseq tool

One of the main disadvantages of quantifying gene expression using RNAseq lies in the large amount of data generated. This requires complex bioinformatics tools and expertise in order to align sequencing reads and analyse the data. This can make the process prohibitive for some studies.

The 3D RNAseq App is an R Shiny app as well as a web-based pipeline for the analysis of any RNAseq data set from any organism. It was developed at the James Hutton Institute and the University of Dundee (Guo *et al.* 2020). This tool allows for the analysis of differential gene and transcript expression, differential alternative splicing as well as differential transcript usage (3D), after mapping of the RNAseq reads obtained from sequencing.

The 3D RNAseq web interface allows for a user-friendly experience of generating gene expression data in a rapid and precise manner, whilst allowing for the handling of complex data sets. It produces several outputs to allow for lab-based researchers to better understand their data, such as heat-maps, volcano plots or gene expression profiles for individual genes.

3.1.3. Aims of the chapter.

In this chapter, I describe how a cheap and simple DNA-methylation assay was developed and optimised with the aim of determining levels of DNA-methylation in plants before and after treatment with a demethylating agent (zebularine, Chapter 4) or carrying mutations in the DNA-methylation gene *met1* (Chapter 5). The method, in common with some of those described above, is based on methylation-sensitive digestion of genomic DNA (gDNA) by the restriction enzyme *HpaII* combined with fluorescent cytosine extension. It was chosen for its capacity to be routinely used whilst not requiring complex equipment and/or high levels of expertise. This method is also relatively cost-efficient by not using sequencing technologies.

In parallel, I performed RNAseq analysis on an RNAseq dataset which was obtained from *Arabidopsis* seedlings treated with 100 µM of the demethylating agent 5-azacytidine (Griffin *et al.* 2016) together with untreated control seedlings. The aim of this analysis was to identify a list of genes that are upregulated, downregulated, or unaffected by the chemical treatment. I then aimed to use a selection of these genes for gene expression analysis in *Arabidopsis* seedlings treated with the zebularine, a 5-azacytidine analogue

with similar demethylating properties (Chapter 4). This gene expression analysis would provide a strategy to characterise the effect of the zebularine treatment on plant development, should changes in DNA methylation be undetectable, either because the assay I developed above did not prove to be sensitive enough, or because the changes were too subtle, or too localised, to be detected.

3.2. Material and Methods

3.2.1. DNA-methylation assay

This assay was first described using DNA extracted from mammalian cell-lines (Zhou *et al.* 2017). Briefly, genomic DNA is submitted to two separate digestions reactions using restriction enzymes *MspI* and *HpaII*. These two enzymes cut the same restriction site (5'-C^CGG-3'). *MspI* is a non-methylation sensitive enzyme and will cut at this site regardless of the methylation context. *HpaII* however, is a methylation sensitive enzyme which will only cut non-methylated restriction sites. Methylation levels of the plants can therefore be evaluated by comparing the cutting activity of *HpaII* to that of *MspI* in different DNA-methylation contexts. This is coupled to a fluorescent cytosine-extension assay, linking the efficiency of each restriction enzyme in different DNA-methylation contexts to levels of fluorescence of the samples.

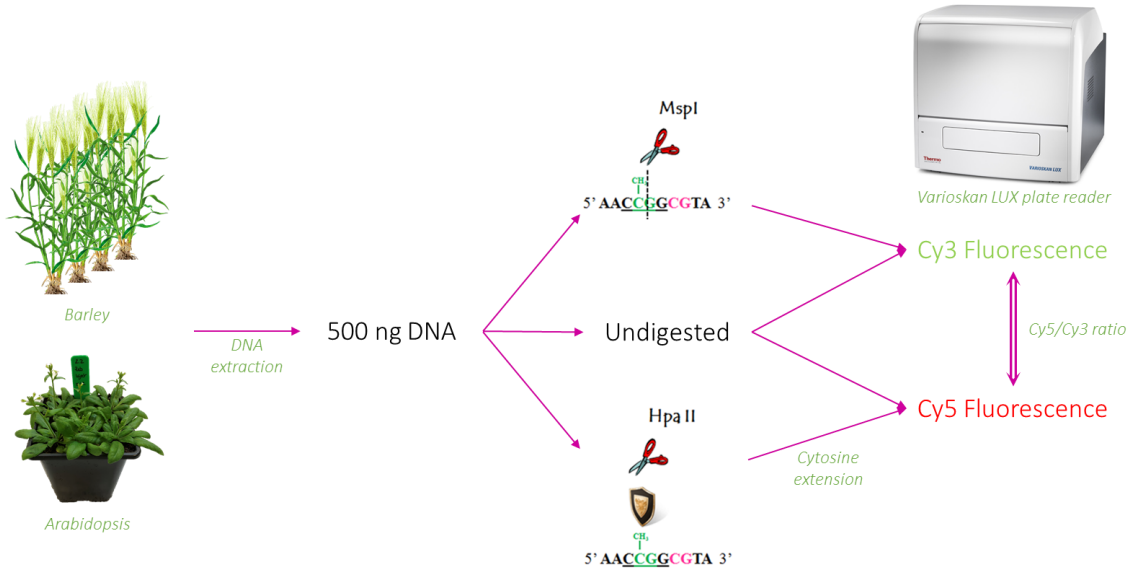


Figure 3.2.1. Overview of the DNA-methylation assay process adapted to plant material for barley and Arabidopsis. Pooled DNA from plants is digested separately with MspI or HpaII, with undigested samples done as negative background controls. The MspI sample and one undigested sample are submitted to cytosine extension using Cy3-dCTP. The HpaII sample and another undigested sample are submitted to cytosine extension using Cy5-dCTP. Difference in fluorescence intensity between the samples is representative of the methylation state of the samples.

Modifications and optimisations were made using this protocol as a model in order to adapt for plant material and guarantee a good enough sensitivity allowing identification of DNA-methylation changes in larger plant genomes such as barley and *Arabidopsis*. Since no bisulfite sequencing data was available on the lines that were used in this project, it was not possible to link the level of fluorescence observed to accurate levels of DNA-methylation. However, this relatively simple and quantitative protocol promised to accurately detect significant differences between methylated and demethylated plants.

3.2.1.1. DNA preparation

DNA extractions were performed as described in **2.2.1** using the maximum amount of leaf tissue possible, 100 mg. This allowed maximum extraction yields to be obtained giving a final concentration of over 50 ng/µL. Samples were normalised by dilution to 50 ng/µL.

3.2.1.2. Enzymatic digestion

Three digestion reactions were prepared including samples containing methylation sensitive enzyme *HpaII*, non-sensitive enzyme *MspI*, and an undigested sample, which was used as a baseline to determine the base level of fluorescence of undigested DNA.

Undigested reactions were composed of 10 µL DNA sample (50 ng/µL), 2.5 µL 10X CutSmart Buffer (New England Biolabs) and 12.5 µL sterile water. Samples which were digested using *MspI* contained 10 µL DNA sample (50 ng/µL), 2.5 µL 10X CutSmart Buffer, 2.5 µL *MspI* restriction enzyme (20 units/µL, New England Biolabs) and 10 µL sterile water. Finally, *HpaII*-digested samples were made of 10 µL DNA sample (50 ng/µL), 2.5 µL 10X CutSmart Buffer, 1 µL *HpaII* methylation-sensitive restriction enzyme (50 units/µL, New England Biolabs) and 11.5 µL sterile water.

Samples were incubated for 30 minutes at 37°C then 30 minutes at 80°C to stop the digestion reaction before being left to cool down to room temperature for a further 30 minutes.

Efficiency of the digestion reaction was verified by migrating 5 µL of the digestion samples on a 1.3% electrophoresis agarose gel. Undigested samples were expected to produce a single high molecular weight band whereas *MspI*-digested samples should produce a smear on the gel. The observed products after digestion by methylation-sensitive enzyme

HpaII differ depending on the methylation state of the DNA from a single high molecular weight band to various topologies of smear.

3.2.1.3. Cytosine extension

Cytosine extensions used two types of fluorescent dCTPs, one coupled to the fluorescent dye Cy3 (λ_{exc} 550 nm, λ_{em} 570 nm) and one coupled to Cy5 (λ_{exc} 649 nm, λ_{em} 670 nm). dCTP-Cy3 was used for cytosine extension of *MspI*-digested samples and undigested samples, whereas dCTP-Cy5 was used for *HpaII*-digested samples and undigested samples.

Cytosine-extension reactions were prepared as follows: 5 µL 5X Colorless GoTaq® Flexi Reaction Buffer (Promega), 1 µL 25 mM MgCl2 (Promega), 0.5 µL 10 µM dGTP (Invitrogen), 0.5 µL 10 µM Cy3- or Cy5-dCTP (Jena Biosciences), 1 µL GoTaq® G2 Flexi DNA Polymerase (5 U/µL, Promega), 10 µL (un-)digested DNA template, 7 µL sterile water.

Samples were incubated for 2 hours at 56°C and then stopped by addition of 4 volumes (100 µL) of high-salt buffer B2 from the PureLink® PCR Purification Kit (Invitrogen).

Samples in buffer B2 were then combined, *MspI*-digested samples with *HpaII*-digested samples, and Cy3-labelled undigested samples with Cy5-labelled undigested samples. Combined samples were purified using the PureLink® PCR Purification Kit (Invitrogen) as described by the manufacturers' protocol without adding more binding buffer B2.

3.2.1.4. Readings

Eluted samples of 50 µL were transferred to black 96-well plates with a clear bottom for fluorescent reading. Fluorescence readings were taken from the bottom of the plate using a Varioskan Lux microplate reader (Thermofisher). Undigested sample readings were used as blanks to obtain accurate readings for both *MspI*- and *HpaII*-digested samples. Readings were exported in table form and *MspI*/*HpaII* ratios were calculated for each biological replicates.

3.2.2. Identifying genes differentially regulated in a demethylated context.

A list of the top 110 upregulated genes in *Arabidopsis* seedlings germinated in the presence of 5-azacytidine, a zebularine analogue had previously been published (Griffin *et al.* 2016). This study also made available the original raw RNAseq reads that were obtained from samples germinated on 100 µM 5-azacytidine and controls. I reanalysed

this latter data set to identify upregulated genes, downregulated genes, and unaffected genes in the presence of 5-azacytidine. A selection of identified genes were then be used for RT-qPCR analysis of plants germinated in the presence of zebularine to determine whether the impact of zebularine on gene expression is similar to that of 5-azacytidine. The non-differentially regulated genes were used as a base line for normalising expression levels between samples and most importantly between control and treated samples.

3.2.2.1. Mapping of RNAseq reads

Raw sequence files were available from the NCBI GEO database under the accession no. #GSE80300. Quality control of the reads was performed using FastQC (version 0.11.8, http://www.bioinformatics.babraham.ac.uk/projects/fastqc). As the reads were of excellent quality, no trimming was necessary for any file. Sequencing reads were aligned to the *Arabidopsis* reference Transcript Dataset 2 (AtRTD2) (Zhang *et al.* 2017) using Salmon (Patro *et al.* 2017).

3.2.2.2. 3D-RNAseq analysis

Differential expression analysis was carried out using the online tool 3D RNA-seq (Guo *et al.* 2020). Batch effects were removed with using the RUVseq (Removing Unwanted Variation in RNAseq) method (Risso *et al.* 2014), which uses independent bioreps instead of negative controls to correct data variation. This was necessary as a batch effect was present in the 4th biorep of the control samples, which contained fewer read counts than the other three biological replicates. Data was normalised using the weighted trimmed mean of M-values (TMM) method. Expression analysis was carried out using 3D RNA-seq analysis by adjusting *p*-values using Benjamini-Hochberg False Discovery Rate (*q*-value), and genes were considered differentially expressed for a $\log_2$-fold-change ($\log_2$-FC) above 2.0 or below -2.0, and for $q \leq 0.05$, similar to what had been done in the original study of Griffin *et al* (2016).

3.3. Results and Discussion

3.3.1. DNA-methylation assay

The two-colour fluorescent cytosine extension assay for the determination of global DNA-Methylation (Zhou *et al.* 2017) was first developed using mammalian cell lines and a Typhoon flatbed laser scanner (GE HealthCare Life Sciences). Therefore, optimisation was

needed in order to use this assay on DNA obtained from plant material, specifically from barley and *Arabidopsis*.

The barley genome is larger than the human genome (5.3 Gbp vs. 3.1 Gbp) (Mayer *et al.* 2012, Lander *et al.* 2001). However, the barley methylome represents a smaller proportion of the genome than in the human methylome (up to 32.6% vs 75%) (Malinowska *et al.* 2020, Tost 2010). The original protocol was developed using cell lines, which had been previously characterised as fully unmethylated or fully methylated, with a range of methylation ratios. This meant the assay was optimised to characterise significant changes in DNA-methylation in human cell lines. In this project, the nature of the changes in plant material was unclear and would include demethylating treatments (as described in Chapter 4) or Single Nucleotide Polymorphism (SNP) mutations with potentially minor effects (as mentioned in Chapter 5). Therefore, this assay, when transferred to plant material, needed to account for potentially subtle changes in DNA methylation. Optimisation of this protocol to the plant materials used in this thesis therefore aimed at determining whether the two-colour fluorescent cytosine-extension assay would allow sufficient resolution to identify what could be subtle changes in DNA-methylation patterns.

A few other modifications were made to the original protocol. The fluorescent cytosines used were Cy3-dCTP and Cy5-dCTP in place of the Alexa Fluor 555-dCTP and 647-dCTP, respectively. These dyes had similar excitation and emission wavelengths as their Alexa equivalents, and were less costly, but presented the risk of being more light-sensitive, having intense fluorescence levels and being less stable in solutions. Therefore, extra care was used when handling the samples and completing the fluorescent readings as fast as possible after cleaning the extension assays. Additionally, a Varioskan LUX plate reader (ThermoFisher) was used to read fluorescence levels instead of a Typhoon flatbed laser scanner (GE HealthCare Life Sciences). The Typhoon flatbed laser produces an image as an output, providing the user with more control and a better understanding of the results. For example, the original protocol highlighted the need to mask out the reflections from the reading plates well walls from the images in order to avoid edge effect. In contrast, the Varioskan LUX only produces numerical values of fluorescence readings to the user, making the handling of the data easier, but also making it impossible

to troubleshoot if unusual results were the consequence of reading errors, autofluorescence, or operator errors.

The first optimisation of this protocol consisted of checking there was a differential digestion activity of restriction enzymes *MspI* and *HpaII.* As described in 3.2.1.2, 500 ng of wild type (WT) barley genomic DNA (gDNA) was digested with *MspI* and *HpaII*, with an

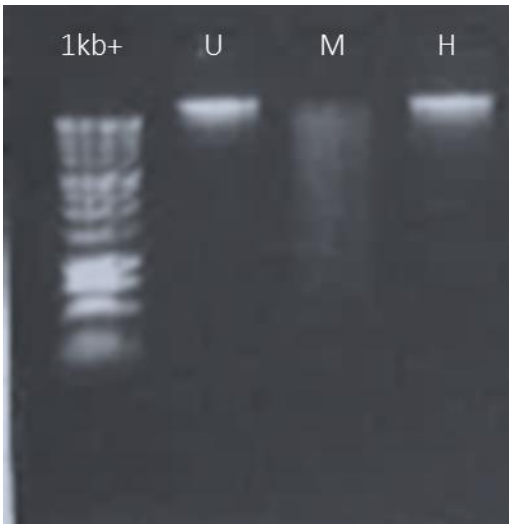


Figure 3.3.1. Restriction digestion profiles of barley genomic DNA undigested (U) or digested with methylation-unsensitive enzyme *MspI* (M) or methylation-sensitive enzyme *HpaII* (H). Samples digested with *MspI* show a smear of digested gDNA whereas samples digested with *HpaII* show a single band, indicating that *HpaII* restriction efficiency was impaired by DNA-methylation on the restriction sites.

undigested sample as a negative control. The digested samples were then analysed on an agarose electrophoresis gel. As can be seen in Figure 3.3.1, the undigested sample resulted in a single clear band of gDNA, suggesting the gDNA was of good enough quality and did not seem to be degraded. Samples digested with *MspI* produced a smear on the gel, indicating the enzyme succeeded in breaking down gDNA in various size fragments. As expected, methylation-sensitive enzyme *HpaII* did not appear to be able to digest the gDNA, which implies that the majority of the *HpaII*/*MspI* restriction sites were methylated. This shows that the activity of *MspI* and *HpaII* is likely dependant on the genomic methylation content, and that the two-colour fluorescent cytosine-extension assay should be able to differentiate between *MspI*- and *HpaII*-associated fluorescence levels.

Next, various 96-well plates used for the fluorescence readings were tested for use on the Varioskan LUX by reading fluorescence levels on empty wells and wells with blank solution (PureLink Elution Buffer, 10 mM Tris-HCl, pH 8.5), at both wavelengths used for Cy3 and Cy5 (550 nm and 649 nm, respectively). Measurements were collected for all

wells of the plates. The type of plate, which showed the least variability between samples was chosen for the DNA-methylation assay. As can be seen in Figure 3.3.2, white coloured plates produced the more variable results, both for Cy3 and Cy5 excitation wavelength. Therefore, white plates were not used for this assay. Two types of black coloured plates were also tested, completely black plates and plates with black sides but a clear base. Additionally, two reading techniques were used for plates with a clear base, top and bottom readings. Bottom readings consisted of a total of 29 separate readings per well, averaged to obtain a single value for each sample. As Figure 3.3.2 highlights, fully black plates showed less variability between samples, both for Cy3 and C5, than black plates with a clear bottom when the readings were taken from above the plate. However, variability was minimised when taking 29 different readings from under the black plates with a clear bottom, again for Cy3 and Cy5 both. For the remaining experiments, black

plates with clear bottoms were therefore used, and readings were taken from under the plate, processing 29 readings for each analysed sample.

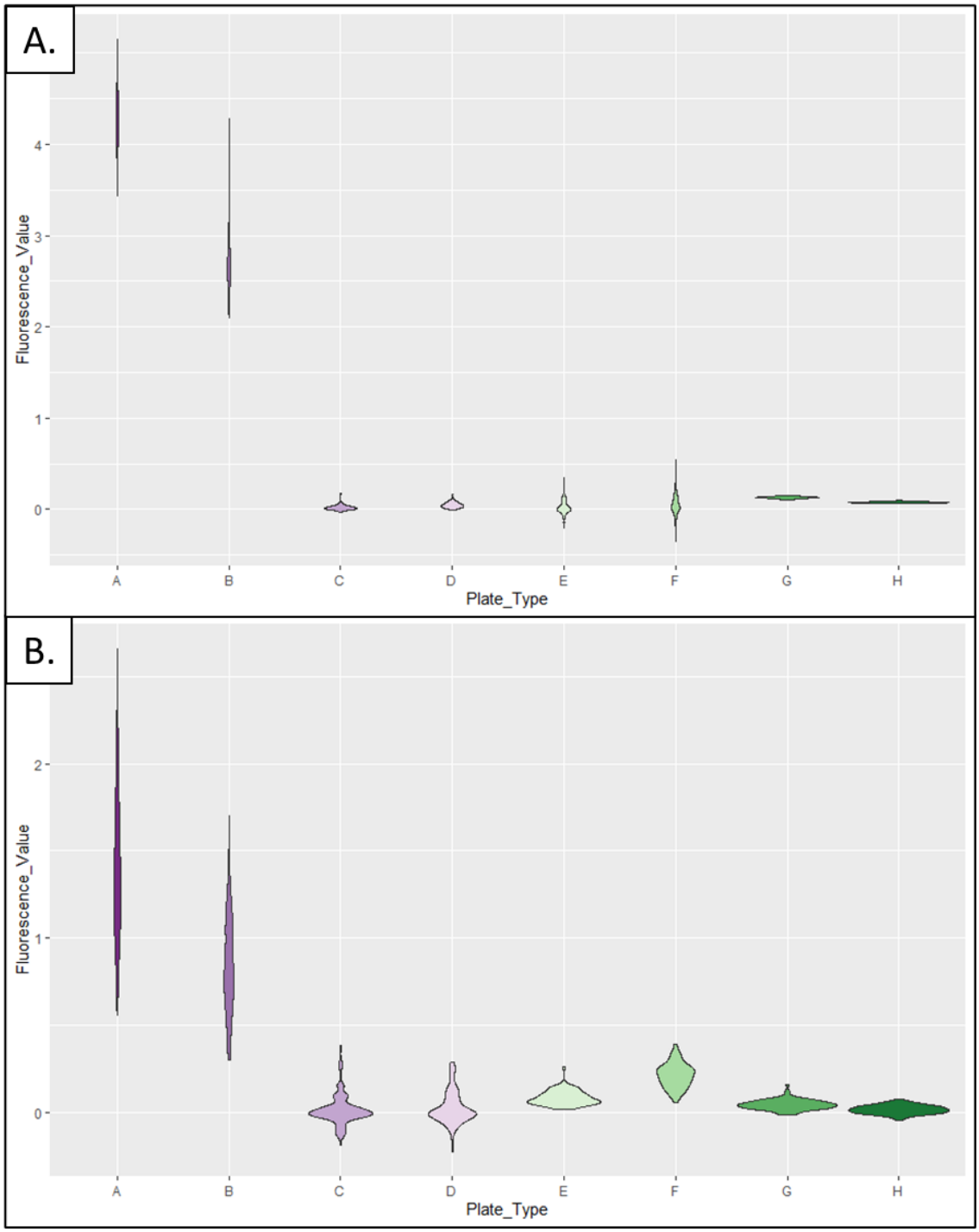


Figure 3.3.2. Comparison of different plate types with readings on the Varioskan LUX for the wavelengths corresponding to the excitation/emission of Cy3 (550 nm/570 nm) and Cy5 (649 nm/670 nm). Top graph A: Readings for Cy3, Bottom graph B: Readings for Cy5. On the X axis, A: empty white plate; B: white plate with 10 mM Tris pH 8.5; C: empty black plate; D: black plate with 10 mM Tris pH 8.5; E: empty black plate with clear bottom, top reading; F: black plate with clear bottom with 10 mM Tris pH 8.5, top reading; G: empty black plate with clear bottom, bottom reading; H: black plate with clear bottom with 10 mM Tris pH 8.5, bottom reading

A standard curve was produced in order to check the quantity of *MspI* enzyme used and that the ratio to input barley gDNA was in excess, resulting in complete digestion of all

available *MspI*/*HpaII* restriction enzyme sites in the genome. This ensured the fluorescence levels observed for *HpaII* can be directly compared to the fluorescence levels obtained for *MspI* as a simple ratio of total available restriction sites, making the assay relatively quantitative when comparing differentially methylated plant genomic material. The same amount of gDNA (500 ng, as described in the original protocol), was digested with increasing quantities of *MspI* restriction enzyme, from 0 to 50 units of enzyme per sample. 50 units correspond to the recommended number of units per sample by the source protocol for both *MspI* and *HpaII*. Three technical replicates were done for each sample, with all conditions emerging from the same pool of gDNA, reducing the risks of biological variance. Figure 3.3.3 shows the fluorescence levels of samples digested with *MspI* and then submitted to Cy3-cytosine extension. Even though this type of graph does not allow for the analysis of the enzyme kinetics in contrast to that of the Michaelis-Menten kinetic model, there seems to be a linear increase of the capacity of *MspI* to digest the same amount of DNA with increasing amounts of enzyme. A plateau is reached around 15 Units of *MspI* per sample, with the enzyme maintaining its efficiency up until 50 units per sample. Therefore, using 50 units of *MspI*, and *HpaII*, per sample, ensures that the enzyme is largely in excess relative to the amount of DNA available.

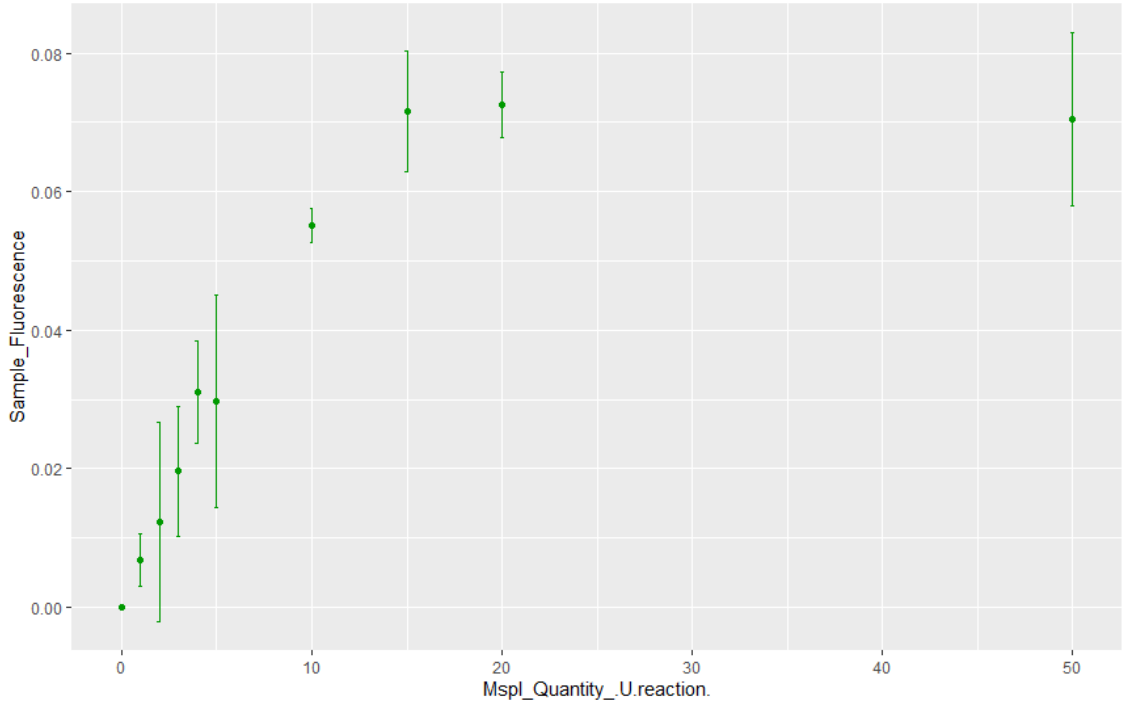


Figure 3.3.3. Standard curve of *MspI* efficiency of identical amounts of DNA for increasing quantities of *MspI* per sample. After digestion, the samples were processed with Cy3-cytosine extension. For each MspI quantity, 3 technical replicates were done using pooled DNA from 5 barley plants.

A proof of concept for the process was performed on gDNA extracted from four different barley plants, following the protocol described in 3.2.1. All plants were 6-week-old wild type plants (cultivar Golden Promise) sown on the same date, gDNA extractions were performed all at the same time. For each plant, three technical replicates were completed in parallel, following the protocol described in 3.2.1. *HpaII*/*MspI* ratios were calculated by removing the undigested samples values from the *MspI* or *HpaII*-digested fluorescence readings and calculating the ratio of the *HpaII*-related fluorescence levels over the *MspI*-related fluorescence levels for each plant and each technical replicate, as explained in Equation 3.3.1.

Equation 3.3.1. Calculations of the *HpaII*/*MspI* fluorescence ratios for the assessment of DNA methylation levels in plants in a relatively quantitative manner. The fluorescence ratio in wild type plants will be used as a baseline to determine the effect of mutations or chemical treatments in other plant material.

$$Fluorescence\ ratio_{\mathrm{HpaII/MspI}} = \frac{\mathrm{HpaII}\ digested_{Cy5} - Undigested_{Cy5}}{\mathrm{MspI}\ digested_{Cy3} - Undigested_{Cy3}}$$

Violin plots were created using the gg_plot2 function in R. As can be seen in Figure 3.3.4, a large amount of variability could be observed between biological samples. This was surprising considering the gDNA was obtained from plants which were grown in similar conditions and of the same age, as well as being wild type plants. No stochastic effects on the levels of DNA-methylation were expected for these biological samples, which could have explained variability between plant samples. Additionally, variability had not been detected when using three technical replicates to obtain the *MspI* digestion efficiency standard curve.

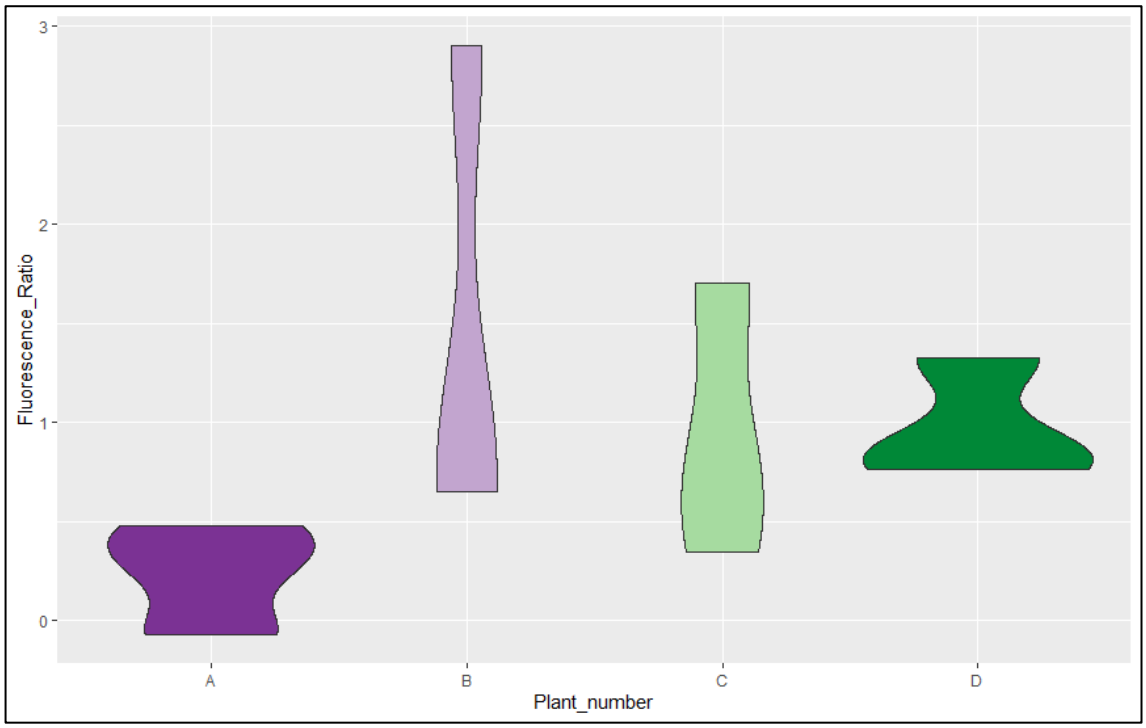


Figure 3.3.4. *HpaII*/*MspI* fluorescence ratios across 4 different plants (A, B, C and D), three technical replicates. Data shows a large variability between plants and technical replicates.

A few options were investigated as to why so much variation could be observed between samples and whether this would impact the capacity and resolution of the assay when looking for potentially subtle changes in DNA-methylation in chemically treated or mutated plants. There was a possibility that the plant material was too old to give a consistent quality of DNA across biological replicates when performing the DNA extraction process. DNA-extractions would normally be done on barley plants at a two-leaf stage, at about 10 to 14 days post-sowing in order to obtain optimal results and good enough quality DNA. As the samples were run in parallel for all of their biological replicates, there was also a risk of user-induced variation between samples and between technical replicates, leading to inconsistencies. After DNA quantification post-extraction, the samples were processed through three different steps: digestion, cytosine extension and purification. All these steps could lead to the amplification of errors from one step to the other and result in different readings between technical and biological replicates. However, whenever was possible, master mixes were used in order to reduce the risks of this error. Furthermore, purification of the cytosine-extension samples was completed in correlation with combining Cy3 and Cy5 samples together, meaning the ratios should not be affected by the purification step. Finally, there is the possibility of error due to the Varioskan LUX plate reader not being fully reliable for accurate readings of fluorescence. Some investigation revealed that previous experiences of colleagues at the James Hutton

Institute indicated that fluorescence and luminescence results could vary strongly between samples within a single plate during the plate reading process, despite routine machine maintenance. The source of the observed inaccuracy of readings remains unclear though it is possibly the result of a sporadic or intermittent fault.

In order to assess whether the Varioskan LUX itself was responsible for the variability in the obtained results, six separate readings were taken using the same sample, in the same plate, at the same well position, without any disturbance of the sample itself. The samples had previously been processed using pooled gDNA of *Arabidopsis* plantlets, with one undigested sample, two samples digested with *MspI,* and two samples digested with *HpaII*. The samples were processed all in parallel following the protocols described in 3.2.1. During the fluorescence reading stage, the plate with the samples was kept either in the plate reader or in aluminium foil to avoid unnecessary light exposure. Readings for undigested samples with Cy3-cytosine extension, which act as a baseline for *MspI*-digested samples, were mostly clustered and do not appear to present much variability (Figure 3.3.5). The standard deviation value for these readings is $\sigma_{\text{Undigested-Cy3}} = 0.004$. Fluorescence readings for *MspI*-digested samples were more variable than the undigested ones for Cy3 fluorescence, but variability between the sample remained relatively low ($\sigma_{MspI\text{-}1} = 0.012$ and $\sigma_{MspI\text{-}2} = 0.009$). However, more variability was observed for undigested samples with Cy5-cytosine extension, which act as a baseline for *HpaII* digested samples ($\sigma_{\text{Undigested-Cy5}} = 0.028$). Similarly, the standard deviation for *HpaII*-digested samples, associated to Cy5 fluorescence readings, was higher for both biological replicates across the six readings on the Varioskan LUX ($\sigma_{HpaII\text{-}1} = 0.028$ and $\sigma_{HpaII\text{-}2} = 0.051$).

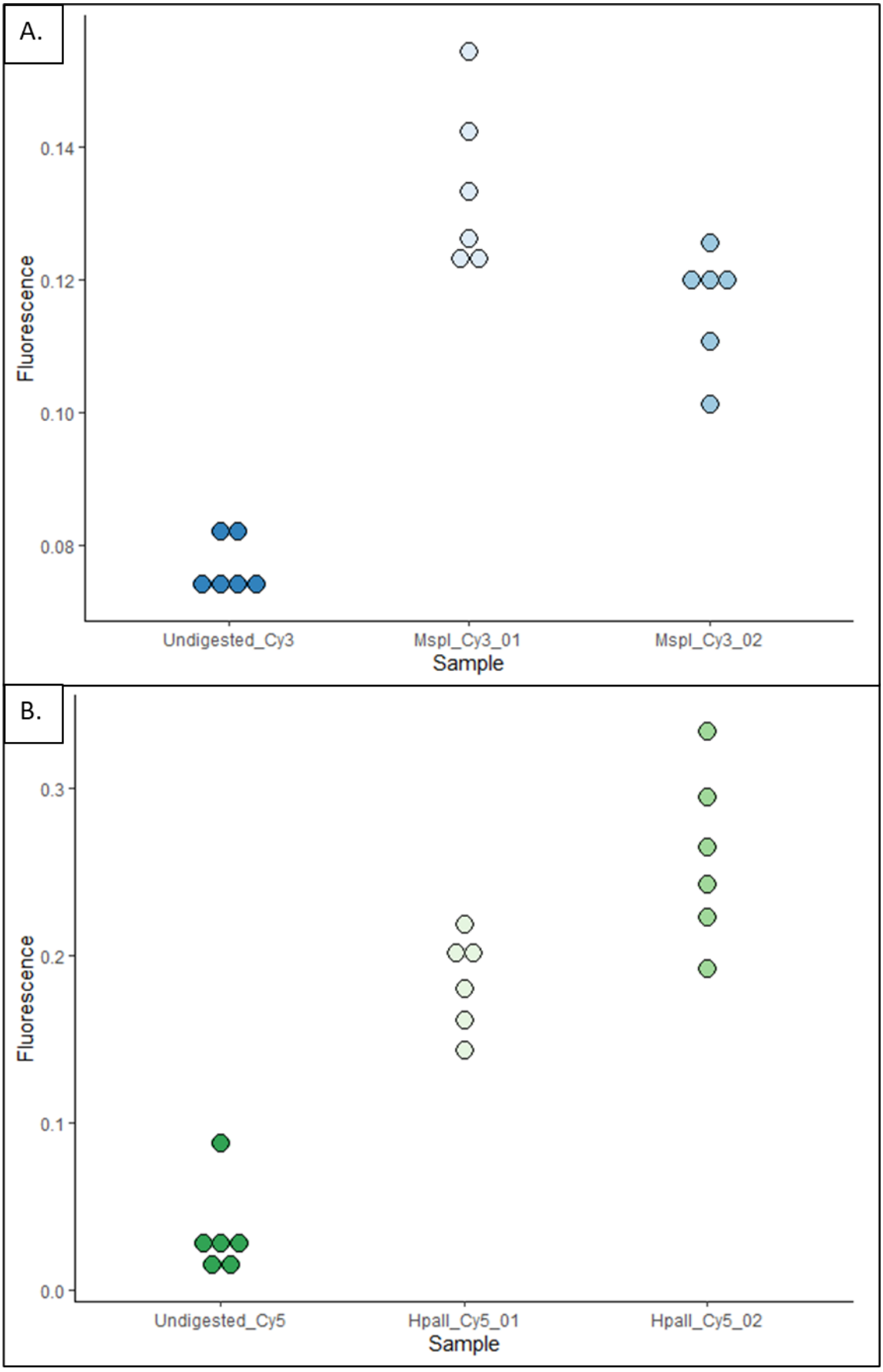


Figure 3.3.5.Cy3 (A.) and Cy5 (B.) fluorescence readings for *Arabidopsis thaliana* gDNA after digestion with *MspI* (A.) and *HpaII* (B.) or undigested (A. and B.). Samples were submitted to fluorescence readings 6 times with no alteration to the sample or the reading plate.

Unfortunately, it was impossible to discern from this data whether the reading inconsistencies are due to a problem of the stability of Cy5 as a fluorophore itself or to incorrect fluorescence measurement at the wavelengths corresponding to Cy5 excitation and emission. It seems very unlikely that the variability between samples was due to a problem with *HpaII* itself as the variability is observed between fluorescence readings of the same sample multiple times with no alteration to the sample itself. It would have been necessary to use another plate reader using Cy5 as a fluorophore or to use samples associated to another fluorophore with similar excitation and emission wavelengths in order to troubleshoot this issue. At the time these optimisation experiments were conducted, no other plate reader was readily available at the James Hutton Institute or the University of Dundee, and in the interest of time and cost, I decided not to test another fluorophore such as Alexa Fluor 647-dCTP.

3.3.2. 3D-RNAseq

The DNA-methylation assay optimisation discussed in 3.3.1 proved to be unable to provide the necessary resolution power to identify subtle changes in chemically treated and mutagenized plants from either barley or *Arabidopsis*. Therefore the effect of treating *Arabidopsis* seedlings with a demethylating agent (zebularine, as discussed in Chapter 4) using gene expression analysis as a reporter of probable hypomethylation of the DNA was assessed instead. Griffin *et al.* (2016) showed that gene expression was altered in *Arabidopsis* seedlings treated with 5-azacytidine and linked these gene expression changes to a global hypomethylation of the genome (Griffin *et al.* 2016). 5-azacytidine is a known analogue of zebularine (Champion *et al.* 2010) and both molecules have been described to similarly demethylate DNA. This means zebularine can be expected to have similar effects on gene expression to 5-azacytidine when used to germinate *Arabidopsis* seedlings. Gene expression analysis of seedlings treated with zebularine could therefore be used as a proxy to estimate the effect of the chemical treatment on global levels of DNA-methylation.

The publicly available RNAseq data associated with the 2016 study of Griffin *et al.* (2016) was used to identify genes that are upregulated, downregulated, and unaffected by a 5-azacytidine treatment using the novel 3D-RNAseq analysis pipeline (Guo *et al.* 2020) as described in 3.2.2.

After mapping reads against the AtRTD2 reference transcript data set, the initial Principal Component Analysis showed that RNAseq data obtained from samples which were exposed to 100 µM 5-azycyticine presented very little variability between biological replicates (Figure 3.3.6.A). In contrast, untreated samples were characterised by a larger amount of variability between samples. This could have been due to differences in sampling strategies or preparation of the samples themselves. Such differences were not mentioned in the original study, however when analysing the number of reads per sample, more variability was observed between untreated, mock samples (Figure 3.3.6.B). Biological replicate number 4 in particular showed a lower read count than the other replicates. To counter this effect, batch effects were removed between biological replicates from the same conditions after normalising the data by calculating counts per millions for each biological replicates. This would allow removal of any variability between samples that is not due to the treatment with 5-azacytidine. A second Principal Component Analysis post-removal of batch effects allowed the removal of a certain level of variability between the samples, even if not completely (Figure 3.3.6.C). The treated samples maintained the low level of variability observed prior to batch effect correction. As can be seen in Figure 3.3.6.D, counts per million variability was reduced between samples, meaning any detected differences between mock samples and 5-azacytidine samples will more likely be due to a biological difference than manipulation or technical variability.

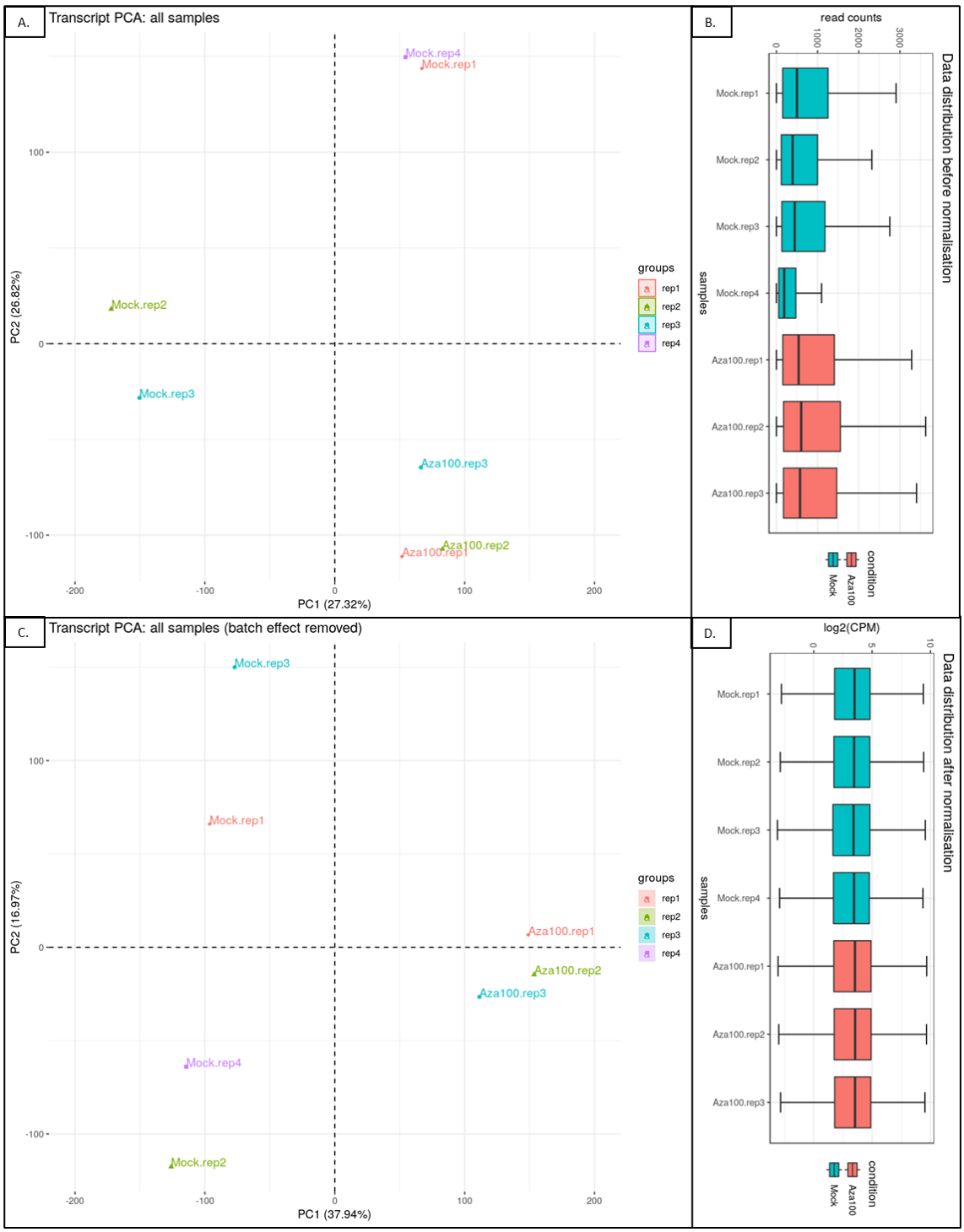


Figure 3.3.6. RNAseq data preparation and analysis of variability between bioreplicates of plants treated with 100 µM 5-azacytidine or untreated. A: PCA plot of raw aligned data; B: read count distribution before normalisation; C: PCA plot after removal of batch effect; D: Distribution of read counts per million reads across samples after removal of batch effect

After differential expression analysis, carried out as described in 3.2.2.2, a heat map was produced to show clear differentiation between mock and treated samples (Figure 3.3.7.A). Unsurprisingly, a large majority of the differentially expressed genes were upregulated in 5-azacytidine-treated samples compared to untreated plants. As discussed in Chapter 1.4, DNA-methylation has been identified as an inhibitor of gene

expression, especially when located in gene promoters. This was further confirmed by a volcano plot correlating expression fold change (FC) for each gene with the statistical significance of the fold change, represented by the False Discovery Rate (FDR) as calculated by the Benjamini-Hochberg statistical procedure (Figure 3.3.7.B). Each individual gene is represented by a data point, with genes significantly differentially expressed represented in red.

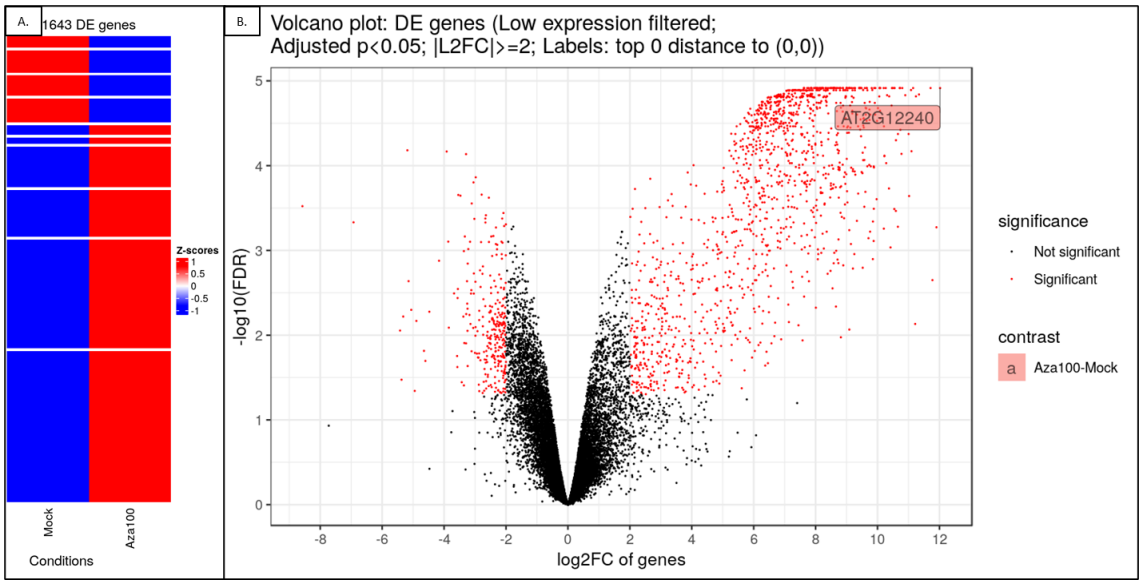


Figure 3.3.7. Differential expression analysis of RNAseq data between *Arabidopsis* seedlings treated with 100 µM 5-azacytidine and untreated seedlings. A: Heatmap showing the repartition of differentially expressed genes, upregulated genes are in red and downregulated genes are in blue; B: Volcano plot showing significantly differentially expressed genes (in red) in the analysed RNAseq data set. Each data point represents a different gene.

Gene expression changes were also verified for genes which had been identified in the 2016 study as overexpressed in the presence of 5-azacytidine and appeared in the top 110 upregulated gene list by comparing their transcript per Million reads (TPM) in both conditions. The most upregulated gene (AT2G11773, hypothetical protein), as well as DUF295 (AT2G17690) and Flowering Wageningen (AT4G25530), which had been studied further in the original study (Griffin *et al.* 2016), were all identified as highly upregulated in plants treated with 5-azacytidine (Figure 3.3.8). This confirmed the 3D-RNA pipeline correlated with the original RNAseq results obtained in 2016 and would allow me to select genes that were upregulated, downregulated, or unaffected by hypomethylation.

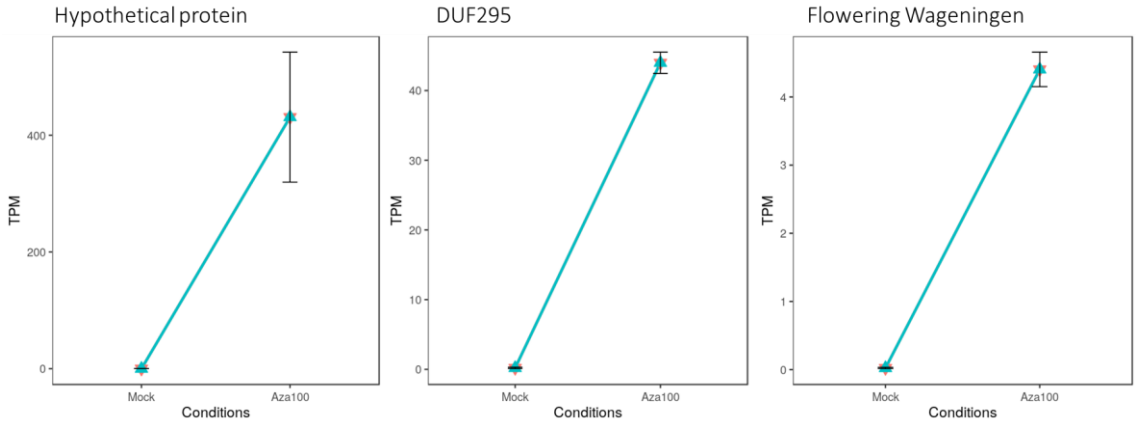


Figure 3.3.8. Gene expression profile for the top upregulated protein as identified in Griffin et al (2016) AT2G11773 (hypothetical protein) as well as DUF295 and Flowering Wageningen, which were also identified as highly upregulated in the initial study. All these genes are correctly identified as upregulated when analysed using the 3D RNAseq pipeline.

This analysis thus allowed me to identify and prioritise 3 upregulated genes, 3 downregulated genes and 3 genes for which gene expression was unchanged in the presence of 5-azacytidine (Table 3.3.1). The three genes with unaffected expression were used as a standard for normalisation across samples when analysing the effect of a demethylating chemical treatment on the other genes, which are predicted to be either up- or down-regulated. I considered it important not to rely on traditional housekeeping genes as references because there was no available information on the effect of DNA-methylation on these genes' levels of expression. The identified upregulated and downregulated genes and their potential role in plant development and stress response will be further discussed in Chapter 4.

Table 3.3.1. List of nine genes which have been identified as unaffected by 5-azacytidine (No effect), upregulated or downregulated. These genes will be used in Chapter 4 to characterise the effect of zebularine on *Arabidopsis* development.

Gene ID	Protein name	Protein function	Effect of 5-azacytidine on expression
AT1G13440	GAPC-2	Glyceraldehyde-3-phosphate dehydrogenase C2	No effect
AT1G04410	C-NAD-MDH1	Cytosolic-NAD-dependent malate dehydrogenase 1	No effect
AT4G34620	SSR16	Small Subunit Ribosomal protein 16	No effect
AT4G16215	N/A	Hypothetical protein	Upregulated
AT2G17690	DUF295	Suppressor of DRM1, DRM2 and CMT3	Upregulated
AT2G11773	N/A	Hypothetical protein	Upregulated
AT5G20630	GER3	*Arabidopsis thaliana* Germin 3	Downregulated
AT2G01520	MLP328	MLP-like protein 328, involved in vegetative to flowering stage transition	Downregulated
AT1G73330	ATDR4	Drought-Repressed 4	Downregulated

3.4 Conclusion

In this chapter, tools and methods were developed to allow a better understanding of the role of DNA-methylation on plant development both in *Arabidopsis* and barley.

First, a protocol based on restriction digestion of genomic DNA was adapted and optimised for use on plant material. The protocol consists in initial parallel digestions of extracted gDNA with restriction enzymes *MspI* and *HpaII*. Both enzymes recognise the palindromic restriction site 5'-C^CGG-3', however, *HpaII* is unable to digest the sequence if the second cytosine is methylated. *MspI*, on the contrary, will be functional on this restriction site regardless of its methylation status. The digestion process is then coupled

to a fluorescent cytosine assay, which labels digested fragments using fluorescent dyes Cy3 (for *MspI*) and Cy5 (for *HpaII*). A quantitative estimate of the levels of cytosine methylation in plants could then be determined using the fluorescence level ratios between *HpaII*-digested and *MspI*-digested samples. Restriction digestion profiles on barley gDNA showed a difference in digestion efficiency between *MspI* and *HpaII*, suggesting a large number of the available restriction sites was methylated in wild type plants. However, when analysing the fluorescence levels of the samples digested with *HpaII* coupled to Cy5 a large variability was observed between Varioskan reads. This meant without considerable further optimisation, potentially involving the use of different instrumentation, this assay would not allow for an accurate measurement of DNA-methylation levels differences between treated and untreated plants with zebularine, or between wild type plants and plants with mutations in *met1*. In the interest of time, it was not possible for me to carry on optimising this protocol. Time permitting, it would have been interesting to assess whether variability could have been reduced if *HpaII*-digested samples were associated to Cy3, similarly to *MspI*, instead of Cy5. This would have meant that fluorescence readings for *MspI*-digested samples and *HpaII*-digested samples would have needed to be done separately instead of within the same sample, which could lead to a higher risk of user error and a longer fluorescence reading time. Alternatively, other fluorescence dyes such as Alexa Fluor dyes could have been tested, which produce more intense fluorescence levels and are more stable than Cy3 and Cy5. Finally, another fluorescence plate reader could have been tested in order to assess whether (as suspected) the observed variability was caused by the Varioskan plate reader itself.

As a mitigation strategy, a pre-existing RNAseq dataset was analysed using the 3D RNAseq web interface in order to identify differences in the transcript abundance of representative genes as a direct result of changes in DNA-methylation. In 2016, Griffin *et al.* (2016) used RNAseq to identify the top 110 upregulated genes in *Arabidopsis* grown the presence of 5-azacytidine. 3D-RNAseq analysis of the original RNA-seq datasets highlighted how genes are affected in the presence of 5-azacytidine and allowed me to select three key genes, which are upregulated, three genes which are downregulated and three genes which are not affected by the chemical treatment. I then made the assumption that these genes would behave in the same manner in the presence of

zebularine, an analogue to 5-azacytidine. They will be used for the characterisation of zebularine treated plants in Chapter 4.

4. The effect of Zebularine on meiotic recombination in plants

4.1. Introduction

As discussed in 1.4, DNA-methylation plays a major role in plant development by regulating gene expression, chromatin structure and genome stability. However, the extent of the role of this epigenetic marker and the mechanisms involved are not yet fully understood. Many tools have been developed in order to understand the role of DNA-methylation on development in plants and the effect of epi-alleles. Several mutant lines are available in *Arabidopsis*, tomato (*Solanum lycopersicum*) and rice (*Oryza sativa*) for methylation genes such as *met1* and *ddm1*, and RNA-directed DNA methylation (RdDM) mutants *drm1*, *drm2* and *rdr2* (Kankel *et al.* 2003, Zemach *et al.* 2013, Corem *et al.* 2018, Liu *et al.* 2015, Hu *et al.* 2014). Some of these mutant lines have been used to conduct studies on the role of DNA-methylation on meiotic recombination, mostly in *Arabidopsis*. However, a lot is still to be uncovered regarding the mechanisms by which DNA-methylation influences the recombination landscapes in plants.

4.1.1. Available DNA-methylation mutants in *Arabidopsis* and their role in meiotic studies

DNA-methyltransferase MET1 mutants *met1-3* and *met1-4* were first used to characterise the role of DNA-methylation post-meiotically in the *Arabidopsis* germline (Saze *et al.* 2003). Knock-out mutations of the MET1 gene led to a very wide diversification of the methylome in the male and female gametes. This was subsequently resolved by re-methylation of the demethylated templates in the zygotes once a functional copy of MET1 was reintroduced. The effects of *met1* mutations on meiotic recombination were studied using the *met1-3* mutant line. In 2012, epigenetic Recombinant Inbred Lines (epiRILs) were generated by crossing *met1-3* mutants with wild type (WT) plants from the same Columbia accession, then inbreeding the subsequent generated lines for 8 generations whilst segregating the mutated *met1-3* allele out of the population (Mirouze *et al.* 2012). The resulting lines therefore carry the WT allele for MET1 but are characterised by differentially methylated genomes. Drastic changes in recombination rates were observed in this study that were dependent upon the

chromosome context. In demethylated plants, an increase in CO numbers was observed in sub-telomeric regions of chromosome 4, whilst the recombination rate was lower in the peri-centromeric region of chromosome 2. However, the global number of CO events across the genome remained unchanged, suggesting a shift in CO location rather than an increased number of successful CO events in the plants. Due to the nature of the epiRILs, such changes were confirmed as dependent on the methylation context of the genome, and not on the presence of the *met1-3* mutation.

Similar results were obtained in *met1-3* mutant plants crossed with the Fluorescent Tagged Line (FTL) CEN3, which contains a marked region that spans the centromere of Chromosome 3 in an interval of 5 Mb/11 cM (Yelina *et al.* 2012). A significant decrease in genetic distance was observed in the F_1 plants obtained from this cross, but a stochastic effect on the CO rates was identified in the F_2 generation, suggesting the effect of the *met1-3* mutation on meiotic recombination was not consistent across generations. Furthermore, *met1-3* mutants were used in combination with *zip4* and *fancm* mutants in order to determine whether hypomethylation of the genome had an impact on interference-dependant CO events (Yelina *et al.* 2015). The total number of CO events in interstitial regions was unaffected in *met1-3/+ zip4* double mutants, suggesting the effect of the *met1-3* mutation is dependent on the interfering repair pathway of CO formation. In contrast, the decrease in CO events observed in *met1-3* mutants in centromeric regions was countered by the effect of the *fancm* mutation in *met1-3/+ fancm* double mutants but did not achieve the levels observed in *fancm* single mutants. These data suggest global hypomethylation of the genome in *met1-3* mutants leads to a redistribution of CO interference across the genome, leading to changes in the recombination landscape in demethylated plants.

Interestingly, an epigenetic activation of meiotic DSBs was observed in *met1-3* mutants when mapping DSB occurrence along the genome (Choi *et al.* 2018). This higher DBS activity in demethylated mutants coincides particularly in centromere regions with a transposon-rich sequence context. This highlights complex links between epigenetic markers such as DNA-methylation, chromatin structure and meiotic DSBs, which have still to be elucidated.

In a similar fashion to the *met1* mutants mentioned above, mutants in *decreased DNA methylation 1* (*ddm1*) were used for meiotic recombination analyses across the genome.

The loss of DDM1 function in *Arabidopsis* led to a significant increase in CO rate in sub-telomeric regions in both heterozygous and homozygous mutant lines (Melamed-Bessudo & Levy 2012). As was observed in *met1-3* mutants, the effects of the *ddm1* mutation varied depending on the chromatin context along the chromosome. More CO events were observed in euchromatic regions, but there were fewer recombination events in heterochromatic regions.

ddm1 mutants were also used to develop epiRILs similar to the lines developed using *met1-3* mutants (Mirouze *et al.* 2012) in order to study the role of DNA-methylation independently of the presence of the *ddm1* mutation in the plants (Colomé-Tatché *et al.* 2012). The variability in methylation patterns between the lines was used as markers of recombination levels and allowed for the creation of a methylation recombination map comprising 126 differentially methylated regions (DMRs) as markers. Analyses using these lines led to further confirmation that pericentromeric regions were subject to additional repression in recombination levels in hypomethylated plants.

In addition to *Met1* and *Ddm1*, which are mainly involved in the maintenance of pre-existing DNA-methylation patterns, the role of genes involved in the RNA-directed DNA-methylation (RdDM) pathway in reproductive cells was also studied in mutants in genes such as Domains Rearranged Methyltransferase 2 (*DRM2*), Chromomethylase 2 (*CMT2*) and *CMT3*. *CMT2* and *CMT3* mainly affect levels of CHG- and CHH-methylation, and mostly do not impact the levels of CG-methylation in somatic tissue (Lindroth *et al.* 2001, Zemach *et al.* 2013). Their role does not vary in cells from the sexual lineage and CG-methylation is not impacted in sperm cells by *drm1drm2* double mutations, nor is it depleted in *cmt2* and *cmt3 Arabidopsis* single mutant plants (Hsieh *et al.* 2016). The use of these RdDM mutants also led to the identification of meiosis specific DNA-methylation islands in meiocytes (Walker *et al.* 2018). These sexual lineage hypermethylated loci (SLHs) are responsible to some extent for the regulation of gene expression in meiotic cells, and depletion of RdDM can lead to abnormal meiotic phenotypes such as disruption of the formation of the spindle in Meiosis I.

Finally, non-CG DNA-methylation patterns were disrupted by the creation of mutants for histone 3 lysine 9 dimethylation (H3K9me2) which is involved in CHG-methylation by interacting with CMT3. Mutants in the H3K9 methyltransferase genes *KYP*/*SUVH4 SUVH5 SUVH6* show an increase in peri-centromeric levels of CO events in parallel to higher

levels of DSB formation in these regions (Underwood *et al.* 2018). This further confirmed a repressive role of non-CG methylation on recombination rates in these regions, as well as the importance of chromatin structure on meiotic recombination patterns.

The main limit to using DNA-methylation mutants such as *met1-3*, *ddm1*, *cmt2* or *cmt3*, is that in *Arabidopsis*, tomato and rice, these mutations are accompanied by high levels of lethality and very low fertility (Kankel *et al.* 2003, Liu *et al.* 2015, Hu *et al.* 2014). This makes the maintenance of such lines a challenge, especially when homozygous mutant plants are needed for crossing purposes with other lines. The process of studying the effect of these mutations is therefore long and fastidious. Additionally, these lines affect the levels of DNA-methylation at a genome-wide level, and it is not possible to target specific regions to be hypomethylated whilst maintaining the rest of the genome at normal levels of methylation. There is therefore a growing interest in developing tools to modify recombination levels transiently and specifically in plants, which would lead to a better understanding of the epigenetic machinery on developmental processes such as meiotic recombination.

4.1.2. Manipulating DNA-methylation in plants

As was described in Chapter 1.3, *de novo* DNA-methylation is regulated by the RNA-directed DNA Methylation (RdDM) pathway, which relies on the use of small interfering RNAs (siRNAs) to direct specific regions of the genome to be methylated. The Virus-Induced Gene Silencing (VIGS) technology has been used to produce siRNAs which specifically target promoters of genes to be silenced which associate to the proteins involved in the non-canonical RdDM pathway and lead to hypermethylation of the promoter and therefore, silencing of its associated gene. For example, siRNAs were delivered into *Arabidopsis* plants targeting meiotic hotspots *3a* and *3b* in the sub-telomeric region of chromosome 3 (Yelina *et al.* 2015). This led to hypermethylation and significant reduction of CO activity in both those hotspots, confirming that DNA-methylation had a regulatory role on meiotic recombination patterns.

DNA-binding proteins such as TAL effectors (TALE), Zinc finger proteins (ZF) and now CRISPR-dCas9 in particular are of interest in plant science for their capability to specifically bind regions of interest in the genome. This capacity has been used in animal genomes to target regions for precise hyper- or hypomethylation by fusing these proteins to proteins involved in the DNA-methylation regulation process, as reviewed by Lei *et al.*

(2018). In plants, so far, only ZF-fusions have been used to target DNA-methylation or demethylation to specific genes. In particular, a ZF-SUVH9 fusion allowed for hypermethylation of the Flowering Wageningen (FWA) promoter in *Arabidopsis*, leading to silencing of the gene (Johnson *et al.* 2014). In contrast, fusion of the ZF targeting complex to the catalytic domain of Ten-Eleven Translocation 1 (TET1-CD), a protein involved in the DNA-demethylation process, restores expression of FWA after silencing (Gallego-Bartolomé *et al.* 2018). To this date, none of these DNA-binding systems has been used to specifically target meiotic hotspots.

These techniques represent precise tools to better understand the role of DNA-methylation on gene expression, transposon regulation or chromatin interactions during plant development. However, some off-target effects have been observed (Gallego‐Bartolomé 2020) and there is a need for further optimisation of the technology to ensure full specificity and efficacy of the targeting. The examples above relied on genetic modification of full plant lines, but there is a potential for transient application of these changes by direct application of the siRNAs or DNA-binding proteins by exogenous RNA treatments (Dubrovina & Kiselev 2019). However, such techniques still require extensive development to be efficiently used and remain costly and time demanding.

In contrast, the use of chemicals to manipulate DNA-methylation in plants allows for routine experiments on the role of DNA-methylation on plant development. Chemical treatments have the advantage in that they can be temporary and can target specific developmental stages of the plants. Moreover, they can trigger effects in a transient manner in contrast to genetic modifications, which are fixed in the lines. The effect of the treatment on plant development can therefore be reversible and allow for the recovery of full seed sets, making the maintenance of the lines easier and less labour-intensive.

5-azacytidine and zebularine are two chemicals used to demethylate globally plant genomes. Both are cytidine analogues and impair the DNA-methylation machinery by covalently binding to methyltransferases, therefore creating a suicide inhibition and preventing the proteins from methylating cytosines in the genome (Champion *et al.* 2010, Gnyszka *et al.* 2013). These molecules were first identified in human medicine and used as chemotherapy agents for treating a range of leukaemia, bladder and bone cancers (Yoo *et al.* 2004, Cheng *et al.* 2004, Flotho *et al.* 2009), but their demethylating capacities were very quickly identified as a major tool for understanding the role of DNA-

methylation in plant development. Even though both chemicals have been used to demethylate plants genomes, zebularine is preferred for plant treatments because of its higher stability in water and reduced risk to mammalian health (Baubec *et al.* 2009). Zebularine and 5-azacytidine are both cytidine analogues which incorporate in the DNA in-lieu of cytosine and form covalent bonds with DNA-methyltransferases, depleting the number of functional enzymatic proteins in the cells. Thusly, zebularine seems to be similarly efficient to 5-azacytidine at comparable concentrations on *Arabidopsis* plants (Griffin *et al.* 2016).

Zebularine was shown to impact levels of DNA-methylation in a dose-dependent manner in a range of plants such as *Arabidopsis*, *Medicago sativa*, tobacco (*Nicotiana tabacum*), rice (*Oryza sativa L. spp. Japonica*), rye (*Secale cereale*) and wheat (*Triticum aestivum*) (Pecinka *et al.* 2009, Baubec *et al.* 2014, Griffin *et al.* 2016, Majerová *et al.* 2011, Du *et al.* 2014, Ma *et al.* 2016, Finnegan *et al.* 2018). However, the effect of zebularine on DNA-methylation in *Arabidopsis* proved to be only transient, with levels of DNA-methylation returning to normal levels a few weeks after the end of the treatment (Baubec *et al.* 2009).

In *Arabidopsis*, the application of zebularine has been shown to trigger activation of a range of otherwise silenced genes such as Flowering Wageningen (FWA) and Suppressor of *drm1 drm2* and *cmt3* (SDC) in early seedling stages (Baubec *et al.* 2009, Griffin *et al.* 2016). Similarly, the stress response Somatic Homologous Recombination (SHR) pathway was also activated in a dose-dependent manner in zebularine-treated *Arabidopsis* seedlings (Pecinka *et al.* 2009). In tobacco, the application of zebularine leads to an increase in telomerase activity, however the length of the telomeres was not impacted in the demethylated cells. This indicates an additional level of control maintained telomere stability in the treated plants (Majerová *et al.* 2011). Similar effects from zebularine treatment on chromatin structure were observed in large-genome cereals. A Zebularine-treated wheat disomic addition line showed an increase in abnormal chromosomal structures in mitotic cells in a dose-dependent manner, which led to the facilitation of insertions, deletions, and translocations of genetic material from alien to wheat chromosomes (Cho *et al.* 2012). Similar structures were also observed in wild type wheat and rye plants, confirming these events were caused by application of zebularine itself (Ma *et al.* 2016). Zebularine treatment of wheat seeds also led to a partial deletion

of the short arm of Chromosome 7B containing the floral promoter *Flowering Locus T1-b* (FT1-b). Plants with such a deletion were characterised by later flowering and an increased number of spikelets (Finnegan *et al.* 2018).

The described effects of zebularine on chromatin interaction in cereals make it a chemical of particular interest to study its role on meiotic recombination, especially as it has been shown to induce DSBs involving homologous recombination repair in mammal cells (Orta *et al.* 2017). Contrary to genetic modification tools, which are heavily regulated and would need to be selected out of the lines by multiple crossing events in order to remove unwanted phenotypes, the application of zebularine could represent a versatile, cost-efficient, and non-labour-intensive way of unlocking genetic diversity in breeding programs.

4.1.3. Objectives of the chapter

In this chapter, zebularine is applied to germinating *Arabidopsis* seedlings in an attempt to replicate previously described *met1* phenotypes (Kankel *et al.* 2003). Fluorescent-Tagged Lines (FTLs) are used as a visual and convenient tool to assess whether zebularine treatments on *Arabidopsis* seeds impact recombination rates in sub-telomeric, interstitial and centromeric regions of the chromosomes using seed-expressed fluorescent markers. As the assessment of DNA-methylation levels, discussed in Chapter 3.3.1, was unlikely to be robust enough to detect an effect of zebularine on the methylome, gene expression analysis was performed on zebularine-treated *Arabidopsis* seedlings by Reverse Transcriptase quantitative PCR (RT-qPCR).

In parallel, the effect of zebularine on meiotic recombination is also assessed in barley (*Hordeum vulgare*) using the commercial F_1 hybrid winter barley Jettoo, kindly provided by Monika Spiller (Syngenta). Recombination rates are calculated using a collection of Single Nucleotide Polymorphism (SNP) markers along Chromosome 6 in F_2 lines derived from F_1 plants treated and not treated with zebularine. Again, gene expression was assessed in zebularine-treated barley seedlings by microarray as an indirect estimate of the effect of zebularine on DNA-methylation.

4.2. Material and Methods

4.2.1. *Arabidopsis thaliana* Fluorescent Tagged Lines (FTLs)

Fluorescent-Tagged Lines (FTLs) were used to measure meiotic recombination frequency in sub-telomeric, interstitial and peri-centromeric regions of the chromosomes (Wu *et al.* 2015). Three Col-0 Traffic Lines (CTLs) 3.9, 2.2 and 4/20 were used, with their reporter intervals described in Figure 4.2.1.

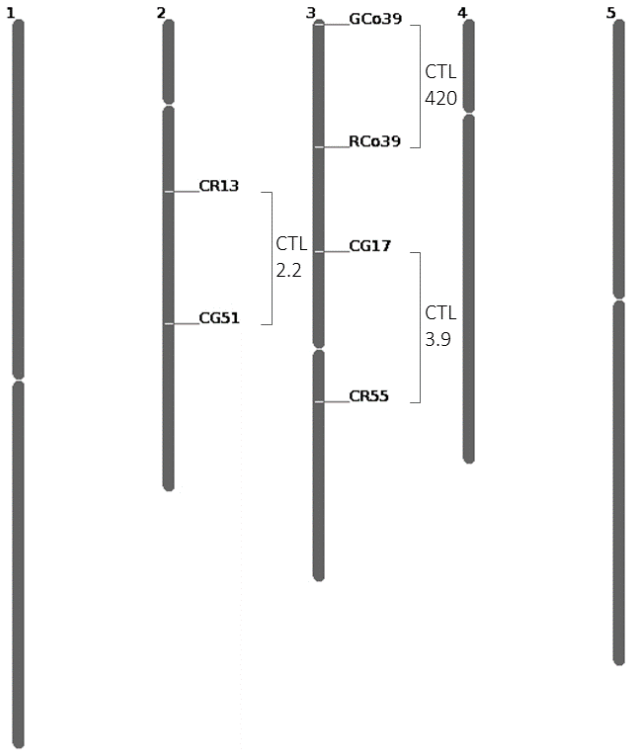


Figure 4.2.1. Three intervals were chosen to measure recombination rates in *Arabidopsis* plants treated with increasing concentrations of zebularine: line CTL 3.9 is used to characterise a 16 cM centromeric interval on chromosome 3; CTL 2.2 is an line with a 21 cM interstitial interval on chromosome 2; CTL 4/20 has a 20 cM interval on the sub-telomeric region of chromosome 3.

These lines, provided by Prof. Ian Henderson (Department of Plant Sciences, University of Cambridge), contain seed-expressed eGFP and dsRed fluorescent proteins regulated by a *NapA* promoter. They allow for measurements of crossover frequency by calculating the number of recombinant seeds for which green and red fluorescent markers have been separated by a recombination event (Figure 4.2.2) as described by Ziolkowski *et al.* (2015).

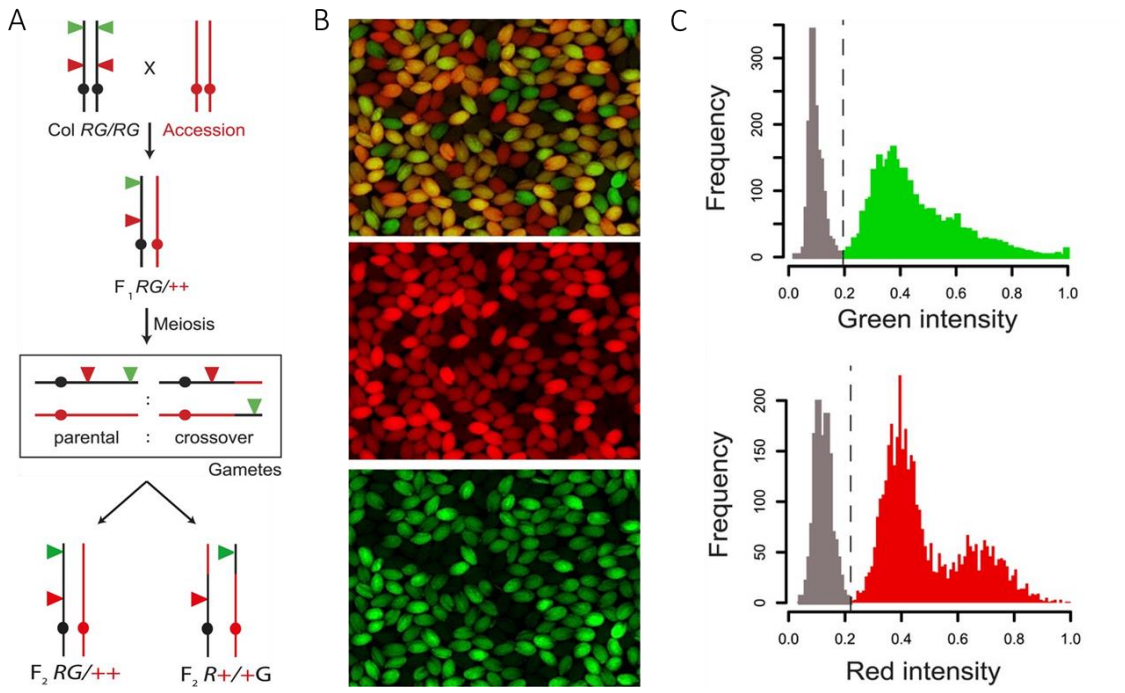


Figure 4.2.2 Use of FTL lines to assess meiotic recombination rate. A: Homozygous FTLs are crossed with a wild-type accession to produce heterozygous F_1 plants. These plants carry both green and red marker genes on the same chromosome inherited from the original FTL. Crossover rate is observed in the F_2 seeds by counting seeds in which the two fluorescent markers have been separated by a CO event. B: F_2 seeds are visualised under red and green filters. When applying a GFP2 filter which will allow for fluorescence of both GFP and dsRed, recombinant seeds appear as only red or only green, whereas non-recombinant seeds are either non-coloured or orange/yellow. C: Using CellProfiler, seeds are separated as coloured and non-coloured depending on their intensity for each fluorescent marker (adapted from Ziolkowski *et al.* (2015).

4.2.2. Sorting and isolating RG/++ F_2 seeds

Ideally, when using FTLs, the number of COs should be counted in F_2 plants produced by F_1 RG/$^{++}$ seeds, which result from a cross between a wild-type line and lines which carry both red and green marker in a homozygous manner. However, given the labour-intense and costly nature of producing F_1 seeds by manual crossing, it is sometimes necessary to select F_2 seeds under the microscope, which appear to be heterozygous for both green and red markers. Seeds were laid under a fluorescence dissecting microscope equipped with GFP3 and mCherry filters (Leica, 470 nm excitation/520 nm emission and 572 nm excitation/635 nm emission, respectively) which allowed for fluorescence of eGFP and dsRed proteins in the seeds, respectively. Seeds were then sorted manually, using a toothpick, first the heterozygous green fluorescent seeds (Figure 4.2.3a) then the heterozygous red seeds (Figure 4.2.3b). Green seeds were sorted first as the resolution of green fluorescence was lower to distinguish between heterozygous and homozygous

and it was easier to sort them when more seeds were present under the microscope, reducing autofluorescence levels.

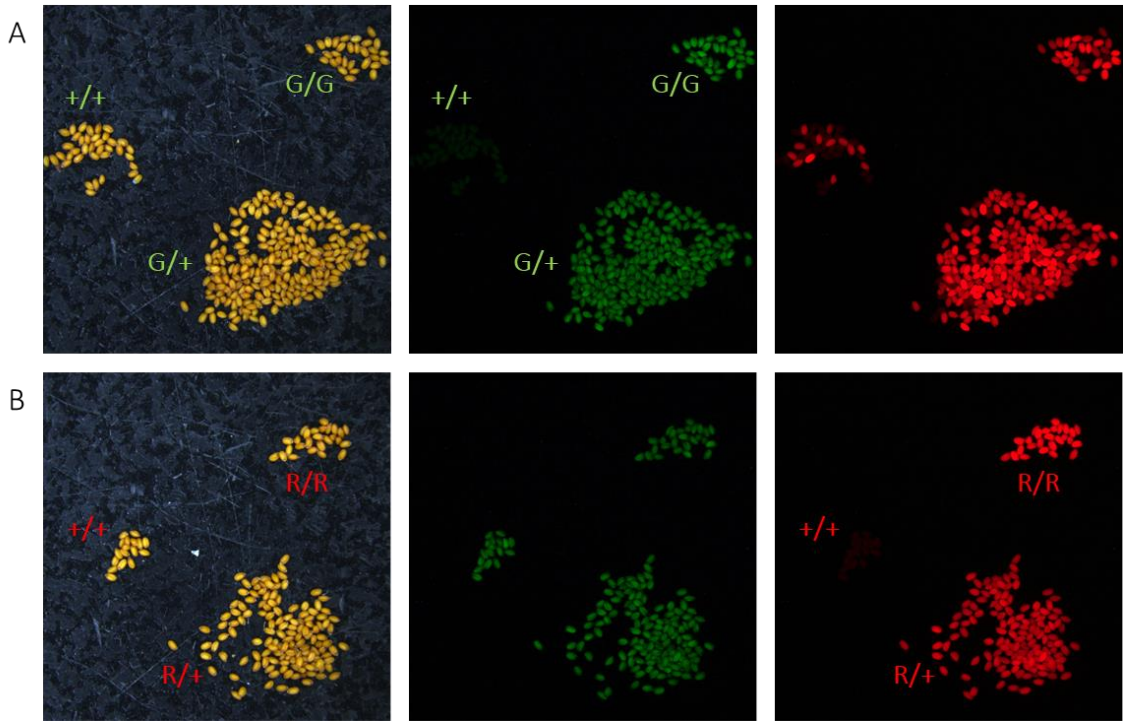


Figure 4.2.3. Sorting process for isolation of RG/$^{++}$ seeds heterozygous for green and red fluorescent markers. A: As it was harder to differentiate heterozygous seeds from homozygous seeds for the green marker, heterozygous green seeds were sorted first. Fluorescence for homozygous green seeds (G/G) was more intense than for heterozygous seeds (G/$^{+}$) whereas non-coloured seeds are not visible under the green filter. B: heterozygous green seeds were then sorted under RFP filters to isolate heterozygous red seeds. Again, fluorescence for homozygous red seeds (R/R) was more intense than for heterozygous seeds (R/$^{+}$). As a result, F_2 RG/$^{++}$ seeds were selected.

Although this strategy avoided having to manually cross the lines carrying the red and green markers with their wild-type counterpart, there was a risk that the selected seeds would not be heterozygous and in coupling. These seeds might also have already been recombinant and have been carrying both mutations in a heterozygous state but on separate chromosomes (in repulsion). Such individuals would however be easy to identify when analysing their offspring as the green/red ratios and expected number of recombinant seeds would be affected and vary from the rest of the population. Visual distinction between the progeny of non-recombinant and recombinant RG/++ hybrids was described previously using a GFP2 filter which allowed the distinction of the different fluorescent classes (Figure 4.2.4, Ziolkowski *et al.* 2015).

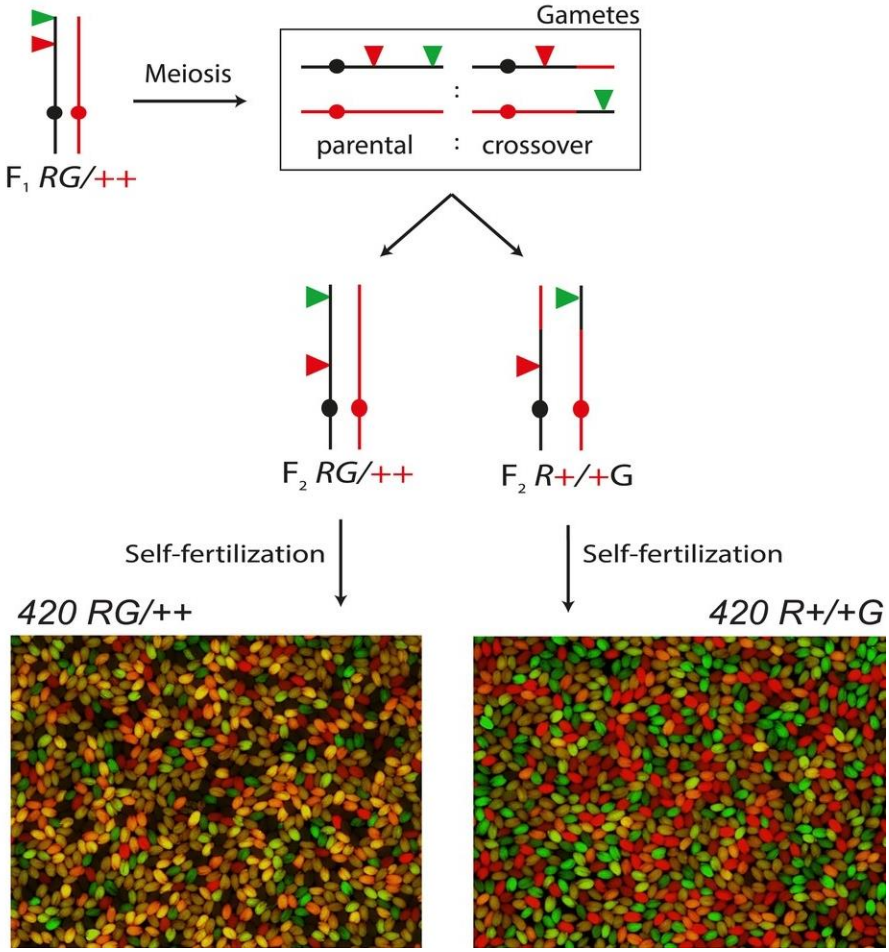


Figure 4.2.4. Distinction between the progeny of true RG/++ seeds and that of already recombined R+/+G seeds was evident as a large proportion of the latter's progeny seed will be carrying only the green markers or only the red marker. This allows for easy identification of seeds obtained from already recombinant F_2 plants and accurate gathering of recombination data (adapted from Ziolkowski *et al.* 2015).

4.2.3. Sterilisation of *Arabidopsis* seeds

Seeds were placed in a 2 mL Eppendorf tube with 1 mL of 10% NaOCl (bleach) and were incubated for 10 minutes whilst being shaken at 1,400 rpm on a Thermomixer (Eppendorf, USA). The tube was then very briefly centrifuged to allow the seeds to set. Under a sterile laminar flow hood, the bleach was removed from the tube and the seeds were washed 5 times as follows: 1 mL of sterile dH_2O was added; the tube was inverted 30 times for mixing; the seeds were left to sit for 30-45 seconds, and the water was then removed. After the final wash, the seeds were resuspended in 0.1% agarose. The seeds were then either used immediately for germination on plates or they could be stored at 4°C for up to 2 weeks.

4.2.4. ½ MS preparation

For *Arabidopsis* seed germination, ½ MS medium was prepared as follows, for 500 mL: 1.075 g of MS powder without vitamins (Model n° M0221 by Duchefa Biochemie); 2.5 g sucrose (Sigma Aldrich); 4 g Agar (Sigma-Aldrich); topped up with sterile water. The medium was then sterilised using a table-top autoclave and let to cool down before being used.

4.2.5. Preparation of DMSO ½ MS plates

Under the laminar flow hood, plates of five different Dimethyl sulfoxide (DMSO, Sigma-Aldrich) concentrations were prepared as follows for 0%, 0.1%, 0.2%, 0.4% and 0.5% final concentration of DMSO: 0 µL, 33 µL, 67 µL, 100 µL, 133 µL, 167 µL of DMSO, respectively, topped up to 30 mL with ½ MS as prepared in 4.2.4. The plates were left open under the laminar flow hood until the medium was solidified to avoid condensation.

4.2.6. Preparation of zebularine ½ MS plates

Plates were prepared for three different concentrations of zebularine under the laminar flow hood, using the previously prepared ½ MS medium (4.2.4) and commercial zebularine (Sigma-Aldrich) resuspended in DMSO (Sigma-Aldrich) to a concentration of 100 mM. Each plate was made of a 30 mL solution prepared as follows for 0 µM, 40 µM and 80 µM plates: 0 µL, 12 µL, 24 µL of zebularine, respectively; 24 µL, 12 µL, 0 µL of DMSO, respectively; topped up to 30 mL with ½ MS medium with 0.5% sucrose (0.8% agar) as prepared in 4.2.4. DMSO was added to the lower concentrations to ensure its concentration was consistent between all the samples, at 0.08%. To avoid condensation, plates were then left open for the medium to solidify under the laminar flow hood.

4.2.7. Germinating *Arabidopsis* seeds on zebularine plates

Under the laminar flow hood, seeds resuspended in 0.1% agarose (4.2.3) were picked up in a 1,000 µL sterile tip with filter using a 1,000 µL pipette, without filling the tip above a third of its capacity. In the tip, the seeds needed to be separated from each other to facilitate placement on the plates. Seeds were then placed one by one on the plate by capillarity as they fell to the bottom of the tip. If two seeds were placed next to each other, one of them was carefully removed using a fresh sterile tip. Once all the seeds were placed on the Petri dishes, these were sealed with two layers of micropore tape.

The plates were then placed at 4°C for 48 hours in the dark before being transferred to a growth cabinet at 20°C (16 hours day, 8 hours night).

4.2.8. Transfer of *Arabidopsis* seedlings from plates to *Arabidopsis*-specific soil

Seedlings were transferred to 9 cm square pots at the Plant Sciences Department at the University of Cambridge. They were grown on a standard commercial soil supplemented with vermiculite as described in 2.1.1. Seedlings were carefully lifted off the medium in the plates using fine tweezers and remaining medium on the roots was removed using distilled water. A small hole was created on top of the soil and the seedlings were placed in that hole using water to push the roots down in the pot. Depending on the size of the seedlings, 6 to 8 plants were placed in each pot. The plants were then left to grow in controlled growth chambers as described in 2.1.1, without the need for a four-day stratification period at 4°C.

4.2.9. Screening seeds for recombinants

Seeds were recovered from fully dry plants for each plant separately. Seeds were then thoroughly cleaned using a fine sieve and white paper to remove all debris and dead seeds. This step was crucial in removing any contaminant, which would be a source of autofluorescence when visualising the seeds under the microscope, and therefore compromise the accuracy of counting recombinant seeds.

Seeds were imaged using a Leica MZ FLIII dissecting epifluorescence microscope (Leica Microsystems) coupled to an Axiocam 506 Color imaging camera (Zeiss) as a single layer spread on a dark surface with the help of a clean glass microscope slide. Approximately 2,000 seeds were imaged for each line in a single image using brightfield light, a GFP filter (excitation 470/40 nm, barrier 525/50 nm) and an RFP filter (excitation 546/10 nm, barrier 590 nm). Images were optimised prior to acquisition using the Zen 3.0 Blue Edition software (Zeiss) by changing the exposure time, gamma values and image temperature in order to maximise the contrast between seeds.

Images for CTL4/20 and CTL3.9 were then enhanced using ImageJ (Schneider *et al.* 2012). Brightness/Contrast of the bright field images was enhanced in order to improve seed outline detection by the seed counting software. dsRed images and eGFP images were treated for red and green colour balance, respectively, which allowed better distinction between coloured and non-coloured seeds by CellProfiler. Settings used for each FTLs are

detailed in Table 4.2.1. Images were then analysed using the software CellProfiler and an in-house analysis script developed and kindly provided by Piotr Wlodzimierz (University of Cambridge, Department of Plant Sciences). This analysis consists in the distinction and counting of coloured and non-coloured seeds for eGFP and dsRed fluorescent markers and the calculation of the number of recombination events in the F_1 FTL seeds for each line.

Table 4.2.1. Image J Brightness/Contrast and colour balance settings used for seed images taken for lines CTL4/20 and CTL2.2. Scales for the values are arbitrarily decided by ImageJ and do not represent an internationally recognised unit.

Line	Brightfield	Red	Green
CTL4/20	0/130	0/165	6/60
CTL2.2	0/130	0/150	0/70

When analysing seed images for CTL3.9, it was observed that the level of fluorescence of the eGFP marker in this line was not sufficient to allow for a clear distinction between coloured and non-coloured green seeds. This meant an automatic counting of recombinant seeds using the Cell Profiler pipeline could not be done. In order to count the recombinant seeds, the images taken with the RFP and GFP filters were overlaid and divided in a 6x8 grid. Twelve cells (25% of the image) were randomly selected for seed counting in a similar fashion to cell-counting when using a haemocytometer: when seeds were counted if they are contained within the square or overlapping the bottom or right-hand borders of the grid but were ignored if they overlapped the top or left-hand borders of the squares. The same twelve cells were used for all plant lines to reduce variability due to lighting differences within the images themselves. The total number of seeds was counted for each square as well as the number of recombinant seeds which were red, but not green, or were green, but not red. The recombination rate for each line was calculated as the ratio of the number of recombinant seeds over the total number of seeds.

4.2.10. DNA Methylation analysis

DNA was extracted from *Arabidopsis* seedlings of accession Col-0 (2.2.1) which were grown on ½ MS with or without zebularine as described in 4.2.7. DNA-methylation was assessed as described in 3.2.1. For each zebularine concentration, five bio-replicates are analysed which contain 80-90 mg of plant tissue prior to DNA extraction. This represents between five and 20 seedlings depending on their size. Each bio-replicate is processed

with three technical replicates and fluorescence readings of each sample were taken six times to account for the variability of the Varioskan plate reader, as discussed in 3.2.1.

For each bio-replicate and each technical replicate, fluorescence readings of undigested samples were used as blanks and removed from the value of the fluorescence readings of their respective digested samples. Ratios were then calculated for each biological and technical replicate between the fluorescence value linked to *MspI* digestion (methylation insensitive) and the fluorescence value associated to *HpaII* digestion (methylation sensitive). Hypomethylated samples would be subject to increased *HpaII* digestion, resulting in higher fluorescence values for *HpaII*-digested samples. This in turn would lead to the *MspI*/*HpaII* fluorescence ratio to be decreased in hypomethylated plants compared to normally methylated individuals.

4.2.11. RT-qPCR for gene expression analysis

4.2.11.1. cDNA preparation from DNase treated RNA.

RNA was extracted and DNase treated from pooled *Arabidopsis* seedlings, three weeks post-germination on ½ MS with or without zebularine, as described in 2.2.4.1. cDNA preparation was carried out immediately after RNA isolation, without thawing of the sample, in order to minimise the risks of RNA degradation. A total of around 10 µg of RNA were used for cDNA preparation for each bio-replicate.

cDNA preparation reactions were prepared on ice by adding 1 µL of 50 µM oligo(dT)$_{20}$ (Invitrogen) and 1 µL of 10 mM dNTP mix (Invitrogen) to 21 µL of each isolated RNA sample at 500 ng/µL. The samples were then submitted to a 5-minute incubation at 65°C before being snap-chilled on ice. The reaction mix was then completed for each sample with 4 µL of 5X First Strand Buffer (Invitrogen), 1 µL of 0.1 mM DTT (Invitrogen), 1 µL of RNaseOUT (Invitrogen) and 1 µL of SuperScript III Reverse Transcriptase (10 U/µL, Invitrogen). The samples were then incubated for 60 minutes at 50°C before being inactivated for 15 minutes at 70°C.

4.2.11.2. Validation of cDNA and RNA samples

For each of the genes selected for gene expression analysis (see 3.3.2), PCR was carried out on both RNA and cDNA samples in order to verify no genomic DNA remained in the RNA samples and cDNA preparation was successful and allowed for amplification of amplicons for each gene. PCR reactions were set up as described in 2.2.2.3 and results

were analysed by agarose gel electrophoresis (2.2.5). A positive control was included for each gene using gDNA extracted following the protocol in 2.2.1.

4.2.11.3. Primer efficiency validation

Efficiency of each primer pair was verified by producing qPCR standard curves for each primer pair, using a pool of cDNA from control bio-replicates as a template.

qPCR reactions were set up as follows: 12.5 µL 2X FastStart Mix (ROX, Roche), 0.25 µL 10 µM corresponding TaqMan probe (Roche), 0.25 µL 20 µM Forward Primer, 0.25 µL 20 µM Reverse Primer, 5 µL cDNA (1:12, 1:48, 1:192, 1:768, 1:3,072 serial dilutions) and sterile water up to 25 µL per reaction. Three technical replicates were processed for each cDNA dilution and three controls with no template were also included for each primer pair. Probes associated to each primer pair are identified in Appendix 8.1.

qPCR reactions were run on a StepOnePlus Real-Time PCR system (Applied Biosystems) with the following program: Initial denaturation at 95°C for 10 minutes, 40 cycles of denaturation at 95°C for 15 seconds and annealing and extension at 60°C for 1 minute, with fluorescence readings taken during this latter stage for each cycle.

Results were analysed using the StepOne Software (Applied Biosystems). For each primer pair, amplification efficiency was calculated by plotting the CT value for each bio-replicate at each cDNA dilution value, creating a standard curve. Primers were validated as having an appropriate amplification efficiency if the values, calculated from the slope of the curve, were between 90% and 110%. These values are relative to what is estimated to be the optimal amplification efficiency slope value if primers were 100% efficient and the PCR amplification behaved as an optimal mathematical model.

4.2.11.4. RT-qPCR analysis

Quantitative PCR was carried out on five bio-replicates of cDNA derived from pooled seedlings, which were germinated with or without zebularine (80 µM). For each studied gene and its associated primer pairs, each bio-replicate was submitted to three technical replicates to ensure consistency of the results. In addition, a plate control, composed of pooled cDNA from untreated samples, was used to ensure consistency between 96-well plates.

qPCR reactions were prepared as followed: 12.5 µL 2X FastStart Mix (ROX, Roche), 0.25 µL 10 µM corresponding TaqMan probe (Roche), 0,25 µL 20 µM Forward Primer,

0.25 µL 20 µM Reverse Primer, 5 µL cDNA (diluted to 1:100) and sterile water up to 25 µL per reaction. qPCR reactions were run on the StepOnePlus Real-Time PCR system (Applied Biosystems) following the program described in 4.2.11.3.

Preliminary analysis of the results was done using the StepOne Software (Applied Biosystems) by plotting Fluorescence levels over the course of the qPCR cycles for each analysed gene separately. These plots ensure results are consistent between technical replicates and the qPCR process was performed correctly.

Statistical analysis of the results was then carried out by calculating Relative Quantification (RQ) for each bio-replicate and for each gene. The RQ value was calculated as $2^{-\Delta\Delta CT}$, where ΔC_T was the C_T value of each technical replicates for each gene of interest compared to the first technical replicate for this gene, and $\Delta\Delta C_T$ was the comparison of the ΔC_T value of each tested gene compared to the endogenous control. Statistical significance of variance in between samples was calculated by analysis of variance (Anova).

4.2.12. Barley Jettoo F_1 hybrid

Seeds of the winter F_1 hybrid variety Jettoo were kindly provided by Syngenta®. Plants were cultivated as described in 2.1.2 however, this variety of winter hybrid requires a vernalisation period 2 weeks after germination. Therefore, the plants were placed in a controlled vernalisation chamber (4 °C, 16 hours day, 8 hours night) for 8 weeks before being returned to a controlled greenhouse environment for the rest of the plant life cycle as described in 2.1.2.

4.2.13. Sterilising barley seeds for germination

Barley seeds were sterilised with a 20% bleach solution supplemented with 0.1% Tween® 20 (Sigma-Aldrich) for 20 minutes. The seeds were then washed five times with sterile water under the lamina flow hoods before being placed onto 2 layers of Whatman® filter paper to drain. The seeds were used immediately after sterilisation to germinate either on agar medium or filter paper.

4.2.14. Germinating barley seeds on Phytagel with zebularine

For barley seed germination, 1% Phytagel™ medium was prepared by adding 10 g Phytagel™ powder (Sigma Aldrich) to 1 L sterile water. The medium was then sterilised using a table-top autoclave and let to cool down before being used.

Zebularine Phytagel™ medium was prepared for 4 different concentrations of zebularine under the laminar flow hood, using the previously prepared Phytagel™ medium and zebularine (Sigma-Aldrich) resuspended in water to a concentration of 100 mM. Each medium was made of a 25 mL solution prepared as follow for 0 µM, 100 µM, 200 µM, and 300 µM plates: 0 µL, 25 µL, 50 µL, 75 µL of zebularine, respectively, topped up to 25 mL with 1% Phytagel™ medium as prepared above. Medium was then distributed in a 48-well sterile cell-culture plate with 500 µL of zebularine medium per well, and the plates were left open under the laminar flow hood for the medium to solidify.

Previously sterilised seeds (4.2.13) were picked up using fine tweezers previously sterilised using ethanol and flame and were placed individually into each well of the 48-well cell-culture plates with medium. The plates were then sealed with Micropore tape and placed at 4°C in the dark for 48 hours. The plates were then transferred to a controlled growth chamber for a 12-day germination (20°C, 16 hours day, 8 hours night).

4.2.15. Transferring barley seeds on soil and growing conditions

After germination on Phytagel™, the seedlings were transferred to 9 cm square pots filled with in house ¼ strength Cereal mix (2.1.2) by creating a small hole in the soil and gently pushing the seedlings roots into the ground. Where necessary, excess Phytagel™ medium was carefully removed from the roots using a pair of clean fine tweezers to avoid rotting of the roots in the soil. Seedlings were then watered carefully to secure them into place in the pots. The plants were then grown in a controlled greenhouse environment as described in 2.1.2.

4.2.16. Germinating barley seeds on filter paper

Seed germination on Whatman filter paper was completed adapting the protocol described by Finnegan *et al.* (2018) for germinating wheat seeds on zebularine medium. Briefly, sterilised seeds (4.2.13) were placed in Petri dishes containing two Whatman® filter paper discs soaked with sterile water. The plates were sealed with Parafilm® tape and placed at 4°C in the dark for 4 days before being transferred to a controlled growth chamber (20°C, 16 hours day, 8 hours night) for 24 hours. The chitted seeds were then transferred under the laminar flow hood onto plates containing two Whatman® filter paper discs soaked with either 7.5 mL of a 300 µM zebularine solution (22.5 µL zebularine (100 mM in DMSO, Sigma Aldrich) in 7.5 mL of sterile water) or 7.5 mL of a control

solution (22.5 µL DMSO (Sigma Aldrich) in 7.5 mL of sterile water). The plates were then sealed with Parafilm® tape and placed again in the controlled growth chamber (20°C, 16 hours day, 8 hours night) for an additional 3 days. The plates were aerated for 5 minutes under the laminar flow hood on the second day of the treatment before being resealed with Parafilm® tape.

4.2.17. Gene expression analysis by microarray in barley

RNA was extracted from barley seedlings (*cv.* Golden Promise) following the method detailed in 2.2.4.2 after being germinated on Whatman® paper filter as described in 4.2.16. Four biological replicates were carried out for each condition, seeds germinated on 300 µM zebularine or treated with sterile distilled water. RNA quality after isolation was assessed by agarose gel electrophoresis (2.2.5) and using a 2100 Bioanalyzer (Agilent).

Microarray processing of the samples was performed by the in-house Genomics service provider at the James Hutton Institute using a custom-made Agilent Microarray for barley gene expression analysis (Morris & Hedley 2019). Results were provided back as a Principal Component Analysis plot and a list of genes for which differences in expression were statistically significant after a volcano analysis (Student's T-test and fold-change cut-off, $p \leq 0.05$).

4.2.18. SNP genotyping on Jettoo F_1 seeds

Jettoo being a commercial F_1 hybrid variety, no genotyping data was publicly available and had to be determined prior to recombination rate studies. DNA was extracted from leaf material of a 2-week-old Jettoo F_1 hybrid seedling as described in 2.2.1. The sample was analysed using the barley 50K Illumina Select SNP chip (Bayer *et al.* 2017) by GeneSeek® (Illumina® UK). 50K genotyping data was visualised using Flapjack (Milne *et al.* 2010) and heterozygous markers for this F_1 hybrid line were identified and mapped on the *cv.* Golden Promise x *cv.* Morex physical and genetic maps (Bayer *et al.* 2017) in order to identify regions poor or rich in heterozygous markers. 24 heterozygous markers were then selected evenly spaced along the physical map of chromosome 6 which were then used for SNP genotyping of F_2 families after zebularine treatment.

4.2.19. SNP genotyping on Jettoo F_2 seeds after zebularine treatment

F_2 seeds were recovered from Jettoo F_1 hybrids plants which had been either germinated on 300 µM zebularine or without zebularine as a control. Leaf material was recovered from these F_2 plants 2 weeks post germination (96 plants derived from zebularine-treated F_1, 96 controls from untreated F_1 plants) and provided to LGC who provided a SNP genotyping service using the 24 SNP markers identified in 4.2.18.

SNP quality was first assessed by visualising allele clustering using the SNPViewer software (LGC). The phase of the putative parental genotypes were then manually determined by segregation analysis and the number of CO events determined calculated for each plant by calculating the number of alternating events between homozygous and heterozygous regions along the chromosome. The average number of CO in between SNP markers along the chromosome was calculated in the same manner across all plants from each condition in order to obtain a representation of the mean number of CO events along chromosome 6.

4.3. Results and Discussion

4.3.1. DMSO above 0.1% represses plant growth in *Arabidopsis.*

DMSO was used in several studies using zebularine to demethylate the genomes of plants such as *Arabidopsis* and wheat (Baubec *et al.* 2009, Griffin *et al.* 2016, Finnegan *et al.* 2018), as it is a known co-solvent which helps penetration of tested drugs into the studied organisms (Zhang *et al.* 2016). However, it was important to determine what the effects on the plants of DMSO itself were since this chemical interacts with cell components and DNA. *Arabidopsis* seeds have been germinated on increased concentrations of DMSO for 21 days as described in 4.2.5 and 4.2.7. After 3 weeks, plants grown on ½ MS without DMSO developed normally in the rosette stage and looked green and healthy (Figure 4.3.1). Plates containing seedlings germinated on 0.1% DMSO and ½ MS were similar to control plates with normal looking healthy rosette-stage plantlets. However, seedlings grown on concentrations of DMSO equal to or superior to 0.2% show significant delays in development. Interestingly, plantlets grown on higher DMSO concentrations at 0.4% and 0.3% carried more leaves and were bigger than seedlings grown on 0.2% and 0.3% DMSO. However, these plants remained largely under-developed compared to control

seedlings and plantlets grown on 0.1% DMSO. This suggested that when treating *Arabidopsis* seeds with Zebularine, the DMSO concentration should remain equal to or less than 0.1%. For the rest of the experiments, the final DMSO concentration was fixed at 0.08% across controls and treated plants, to ensure the effect of DMSO itself on plant development would remain negligible.

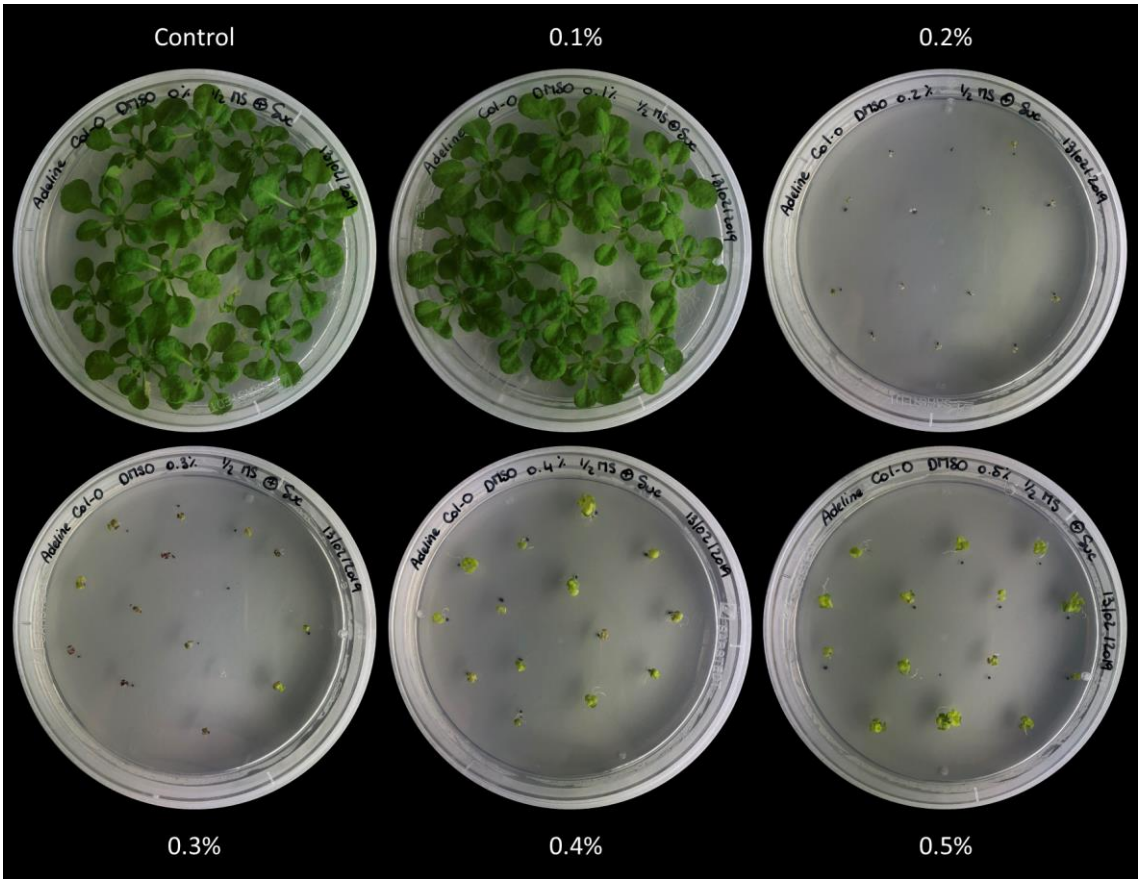


Figure 4.3.1. Germination of *Arabidopsis* Col-0 seedlings on ½ MS with increasing concentrations of DMSO shows a strong effect of larger concentration of DMSO (≤0.2%) on plant development. Photographs were taken at 21 days after transfer to a controlled growth cabinet (20°C, 16 hours day, 8 hours night).

4.3.2. Zebularine causes delayed development in *Arabidopsis.*

Sorted RG/++ F_2 seeds were germinated on ½ MS medium containing increasing concentrations of zebularine for 21 days (4.2.7) for each FTLs used in this project. Control plants grew normally independently of the line, with bright green plants carrying 10-12 leaves each (Figure 4.3.2). However, a dose-dependent response can be observed in plantlets exposed to increased concentrations of zebularine. Seedlings grown on 40 µM zebularine are significantly smaller than control plants with only 8-10 leaves appearing. Similarly, plantlets subjected to 80 µM of zebularine are much smaller than the controls and 40 µM plants, and they only carry 4-6 leaves per seedling. Moreover, plants treated with zebularine are characterised by the appearance of darker, almost black leaves, and

much smaller roots than the controls. Again, these phenotypes were observed independently from the FTLs used.

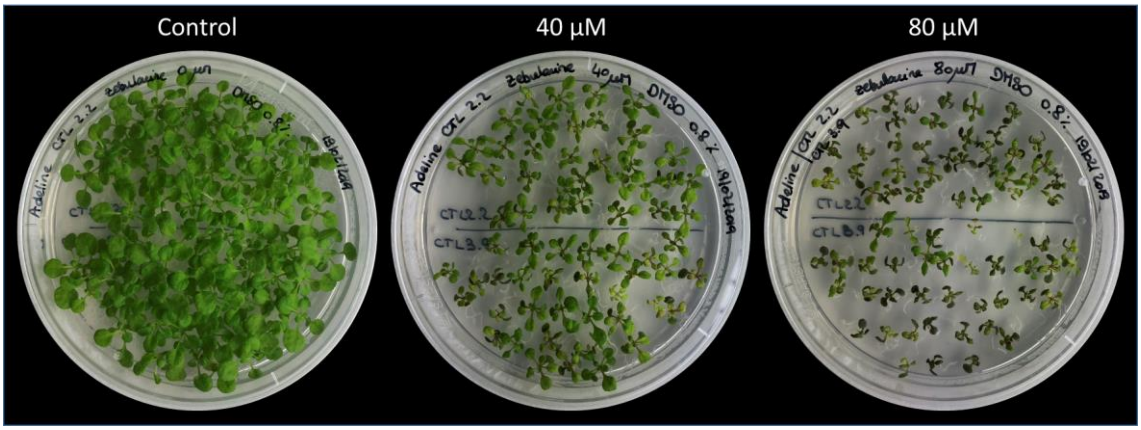


Figure 4.3.2. Dose-dependent response from *Arabidopsis thaliana* seedlings to the application of increasing concentrations of zebularine. Treated plantlets are smaller, with shorter roots, and some of their leaves are darkened.

Seedlings germinated on ½ MS and increasing concentrations of zebularine were transferred into soil as described in 4.2.8. After 5 weeks, plants which had been germinated on zebularine appeared thinner than untreated plants and carried fewer stems (Figure 4.3.3). Treated plants also flowered later than untreated plants and produced fewer siliques. This indicated the effect of zebularine on plant development was strong enough to maintain a delay in plant development and suggested zebularine was still active and having an effect during the meiotic stages and potentially effect changes in the recombination landscape.

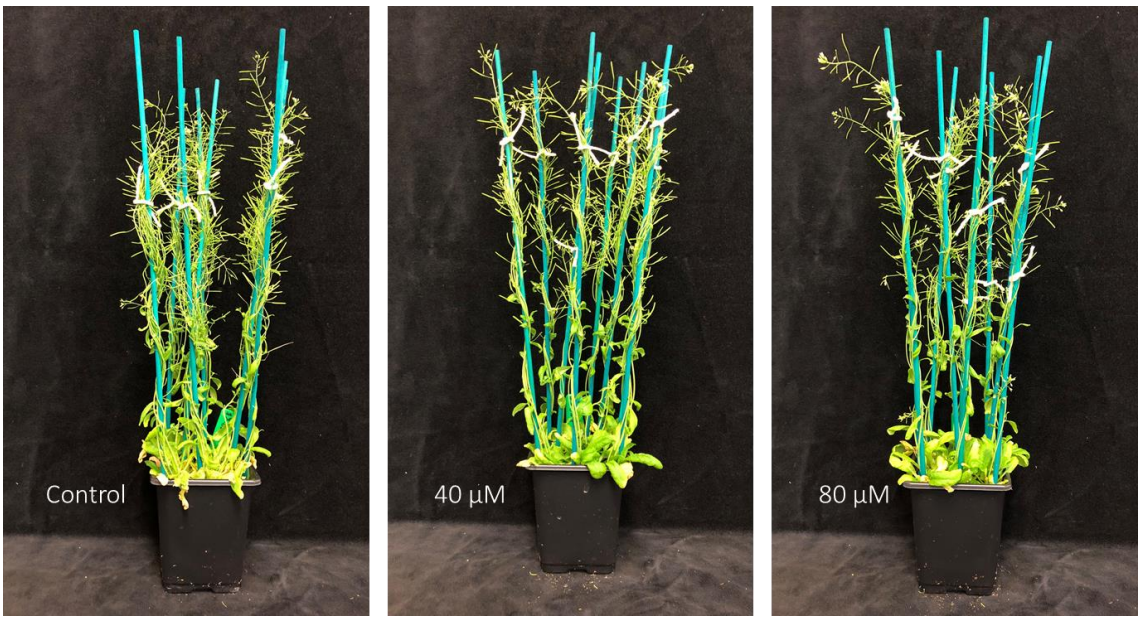


Figure 4.3.3. Five weeks after transfer on soil, plants treated with zebularine still exhibit a delayed development compared to untreated plants. The plants are smaller, produce fewer stems and flower later than the control plants.

4.3.3. Levels of DNA-methylation in *Arabidopsis* seedlings

DNA-methylation levels were assessed on seedlings which were germinated for 4 weeks on ½ MS agar medium with or without zebularine (80 µM, 0.08% DMSO) as described in 4.2.7. Methylation levels were determined using the two-colour cytosine extension assay detailed in 3.2.1 as explained in 4.2.10. As can be seen on Figure 4.3.4, there was no significant difference detected in the levels of *MspI*/*HpaII* fluorescence ratios between *Arabidopsis* plants that were treated with 80 µM zebularine and untreated plants (p = 0.198).

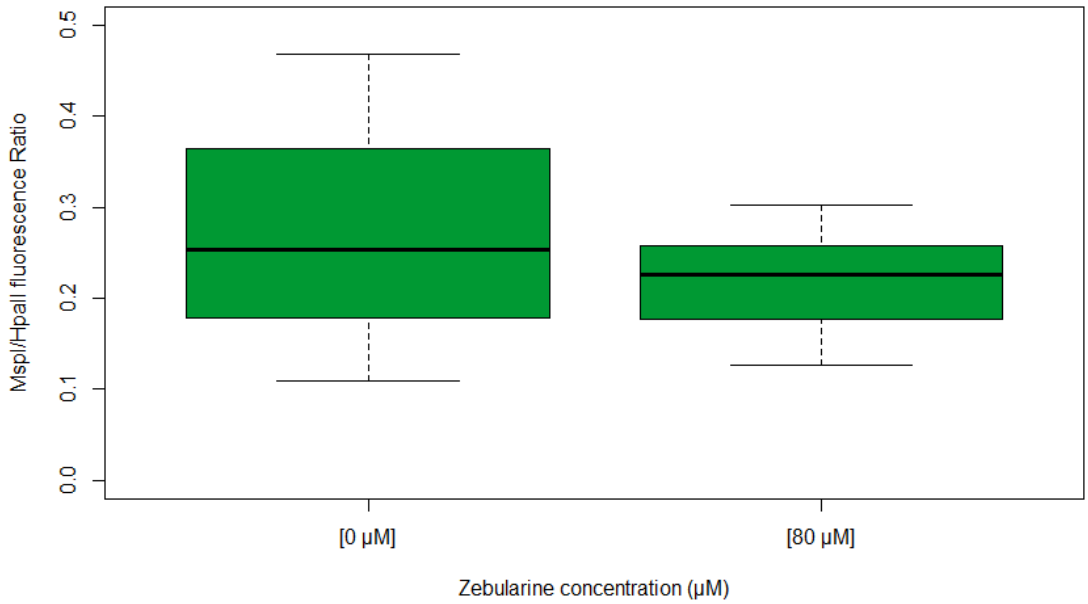


Figure 4.3.4. Comparison of the *MspI*/*HpaII* genomic DNA digestion ratio in plants treated (80 µM) or untreated (0 µM) with zebularine during germination. Difference between the samples is not significant (ANOVA, p = 0.198, n = 5).

Several studies in the past have used Bisulfite Sequencing and High-Performance Liquid Chromatography (HPLC) to determine the levels of DNA-methylation in *Arabidopsis*, rice and wheat treated with zebularine (Baubec *et al.* 2009, Griffin *et al.* 2016, Finnegan *et al.* 2018). All showed a global hypomethylation of the genome in the presence of zebularine compared to control untreated plants. This was expected to be observed when using the newly developed restriction enzyme-based protocol. However, as in 3.3.1, this DNA-methylation assay proved not to be robust enough to confidently identify differences between seedlings treated and untreated with zebularine, despite strong phenotypes suggesting the effect of zebularine on these plants did lead to heavy demethylation in seedlings exposed to zebularine.

4.3.4. Gene expression analysis in zebularine-treated *Arabidopsis* seedlings

As the two-colour fluorescent assay did not allow determination of whether the zebularine treatment did reduce the levels of DNA-methylation in *Arabidopsis* seedlings, I decided to assess whether the zebularine treatment impacted gene expression levels in the treated plants. If gene expression levels are indeed impacted by the chemical treatment, it could be assumed that this is due to a global demethylation of the genome leading to differential regulation of gene expression.

As discussed in 3.3.2, gene expression patterns of traditional housekeeping genes such as elongation factor 1α (EF-1α), ribosomal RNA 18S (18S rRNA) and actin (ACT) were not used as controls in plants treated with a demethylating agent such as zebularine. For this reason, one gene with gene expression unaltered by exposure to 5-azacytidine was selected from RNAseq data on seedlings treated with 5-azacytidine, an analogue of zebularine (Griffin *et al.* 2016). The gene AT1G13440 encodes for a Glyceraldehyde-3-phosphate dehydrogenase C2 protein (GAPC-2) and was chosen as a replacement "housekeeping" gene with a baseline expression level. It is also a close homologue of GADPH, which is traditionally used as a housekeeping gene in RT-qPCR studies. This gene was used in the gene expression analysis to normalise differences between biological replicates and to ensure differences between treated and untreated plants are due to the zebularine treatments and not variances in sampling and cDNA library preparation. Another three genes were upregulated in the presence of 5-azacytidine and three were downregulated in treated plants. These genes were assumed to reproduce the same pattern in plants treated with the analogue of 5-azacytidine, zebularine.

Primer efficiency was assessed for the primers designed to perform RT-qPCR for these 7 genes (as described in 4.2.11.3). Results are summarised in Table 4.3.1 which details the coefficient of determination of each primers' standard curve and their associated efficiency rate. Graphical evidence of the standard curves can be found in Appendix 8.2. Ideally, primer efficiency should be between 90% and 110%, with 100% being the ideal efficiency where the amount of DNA material in the sample doubles every cycle. Having primers with consistent efficiency across genes ensures the differences between genes is due to biological differences between treated and untreated plants and are not artifacts caused by differences between primer pairs. All tested primer pairs amplified correctly, and their efficiency rate was comprised between 89.8% and 112.4%. All these primers were therefore validated as permitting the correct assessment of their corresponding gene's expression level in zebularine-treated plants.

Table 4.3.1. Primer efficiency tests

Gene	Predicted effect of zebularine application	R^2	Efficiency Rate
AT1G13440	No effect	0.991	89.922%
AT4G16215	Upregulated	0.998	91.823%
AT2G17690	Upregulated	0.979	97.354%
AT2G11773	Upregulated	0.983	112.445%
AT5G20630	Downregulated	0.991	89.888%
AT2G01520	Downregulated	0.995	109.018%
AT1G73330	Downregulated	0.992	99.408%

The effect of zebularine application on *Arabidopsis* seedlings on gene expression was assessed by calculating the mean fold change for the 6 genes that were identified in 3.3.2. These changes in gene expression, which are relative to the expression of GAPC-2 (AT1G13440), are summarised in Table 4.3.2. Graphical visualisation of the data is also available in Appendix 8.3.

Table 4.3.2. Gene expression changes in *Arabidopsis* treated seedlings. The chemical treatment caused a statistically significant increase in gene expression for AT2G17690, AT2G11773 and AT5G20630 and a similarly significant decrease in expression for AT5G20630, AT2G01520 and AT1G73330

Gene	Protein	Mean fold change	Effect of zebularine application	*p* value
AT4G16215	Hypothetical protein	127.87	Upregulated	2.22E-04
AT2G17690	Suppressor of DRM1, DRM2 and CMT3	41.24	Upregulated	7.89E-05
AT2G11773	Hypothetical protein	232.30	Upregulated	1.49E-05
AT5G20630	Arabidopsis thaliana Germin 3	0.11	Downregulated	3.84E-03
AT2G01520	MLP-like protein 328, involved in vegetative to flowering stage transition	0.02	Downregulated	1.88E-05
AT1G73330	Drought-Repressed 4	0.03	Downregulated	1.54E-05

This data is consistent with observations in *Arabidopsis* seedlings treated with zebularine analogue 5-azacytidine (Griffin *et al.* 2016). The RNAseq data from that study indeed showed that these genes were upregulated (AT2G17690, AT2G11773 and AT5G20630) or downregulated (AT5G20630, AT2G01520 and AT1G73330) in the presence of 5-

azacytidine. In the original study, the levels of DNA-methylation in plants treated with 5-azacytidine had also been assessed by Whole-Genome Bisulfite Sequencing (WGBS). In plants treated with 100 µM 5-azacytidine, global DNA-methylation levels were significantly decreased, with a drop of CG-methylation from about 17% in untreated plants to around 12% in seedlings exposed to the chemical. Similar effects were also observed in zebularine-treated plants in that same study.

The gene expression analysis in zebularine-treated *Arabidopsis* plants for this project shows similar effects to what had been previously observed in 5-azacytidine-treated seedlings. Despite the difficulties in assessing the levels of genomic DNA-methylation in plantlets treated with zebularine mentioned in 4.3.3, the gene expression data obtained during this experiment suggests that the genome of these seedlings was globally de-methylated. This means any potential effect on meiotic recombination that could be observed in zebularine-treated FTLs seeds will very likely be due to a global hypomethylation of the genome.

4.3.5. Recombination analysis in F_2 FTL seeds

In order to determine whether the application of zebularine would have an impact on meiotic recombination in *Arabidopsis*, three CTLs were used which spanned a sub-telomeric interval (CTL 4/20, 20 cM), an interstitial interval (CTL 2.2, 21 cM) and a centromeric interval (CTL 3.9, 16 cM), as described in 4.2.1. The CTL 4/20 interval was chosen specifically as it was the interval that had previously been used in the 2012 study which showed *ddm1* mutants were characterised by an increase in CO events in this locus (Melamed-Bessudo & Levy 2012). A CO number increase would therefore be expected within the CTL 4/20 locus in F_2 seeds obtained from *Arabidopsis* plants treated with 80 µM zebularine. Likewise, CTL 3.9 is a seed-based reporter line equivalent to CEN3, which is a pollen-based FTL used in the 2012 study which showed an increase in CO numbers in *met1* mutants in the centromeric regions, and eventually led to the identification of crossover hotspot *3a* which is more active in *met1* mutants (Yelina *et al.* 2012). Based on the results from this study, I hypothesised that an increase in the number of CO would be observed in the CTL 3.9 locus in the F_2 progeny of lines treated with zebularine. The interstitial line CTL 2.2 or an equivalent has not directly been used in similar studies using methylation mutants, however based on the results of the 2012 study, which showed a depletion of recombination events using genotyping by

sequencing (GBS) in peri-centromeric regions in *met1* mutants (Yelina *et al.* 2012), it is expected that a decrease in CO number would be observed in the progeny seeds of plants treated with zebularine compared to untreated plants.

Surprisingly, the number of CO events in the CTL4/20 interval was 7.62% lower in the progeny of plants treated with 80 µM compared to control untreated plants (Figure 4.3.5). This reduction is statistically significant (ANOVA test, p = 0.0326).

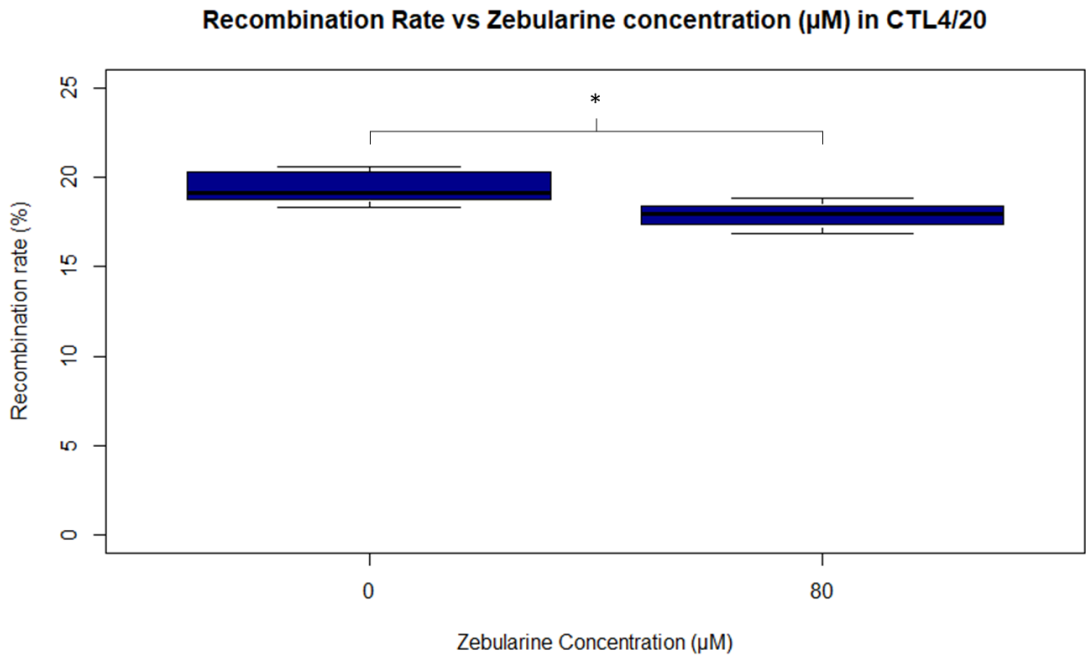


Figure 4.3.5. Zebularine-treated plants show a lower number of CO events in sub-telomeric locus 4/20 (ANOVA, p = 0.0326) compared to untreated plants (n = 6 control, 4 treated, ≤2,000 seeds per plant)

This suggests that despite phenocopying *met1* mutants in the early stages, and in contrast to what has been previously observed in *ddm1* mutants (Melamed-Bessudo & Levy 2012), the application of zebularine to *Arabidopsis* seedlings leads to a CO inhibition in sub-telomeric regions. It had previously been hypothesised that more crossover events were happening in demethylated regions in wild type plants because chromatin void of DNA-methylation would be more accessible to the various protein complexes involved in meiotic recombination (Melamed-Bessudo & Levy 2012). Therefore, it seems unlikely that the decreases in CO events in regions which are traditionally enriched in COs are linked to changes in chromatin state. However, the difference in the mode of action of zebularine, compared to *met1*, at the gene control level, is yet to be fully characterised. It is therefore possible that the meiotic recombination machinery itself was impaired by the lower levels of methylation.

Recombination levels in the interstitial region CTL2.2 of chromosome 2.2 were 1.94% lower in treated plants (80 µM) compared to control untreated *Arabidopsis* seedlings (Figure 4.3.6). This difference was however not significant (p = 0.383).

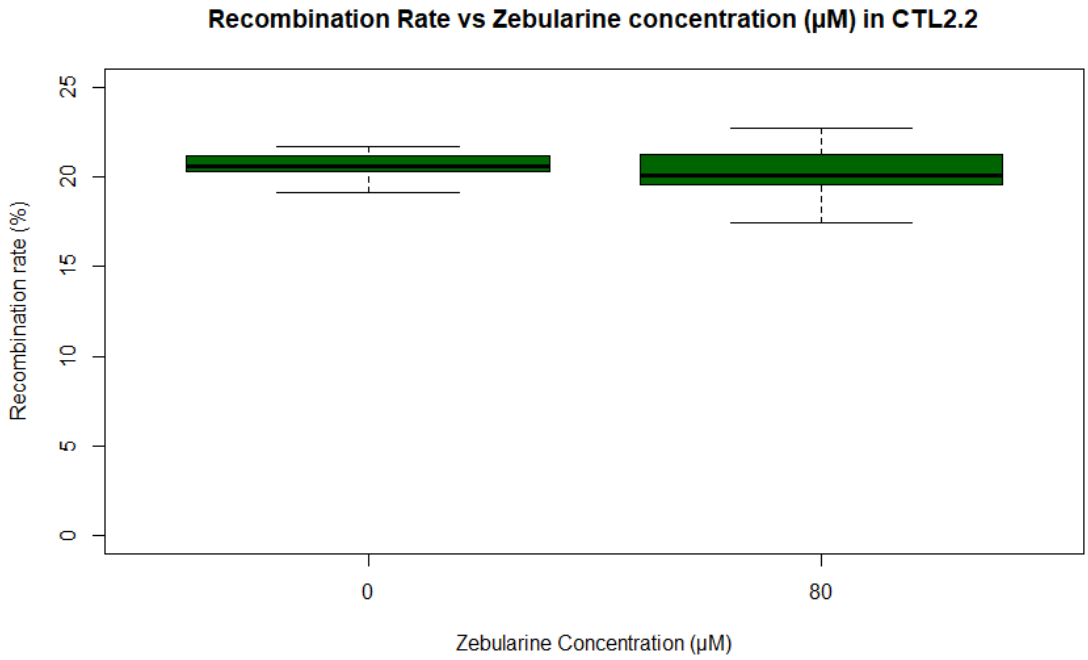


Figure 4.3.6. Zebularine-treated plants show a lower number of CO events in interstitial locus 2.2 (Anova, p = 0.383) compared to untreated plants (n = 14 control, 14 treated, ≤2,000 seeds per plant)

These observations are consistent with previous data described in *met1-3 Arabidopsis* demethylated mutants in peri-centromeric regions (Yelina *et al.* 2012). In that study, the effect of the homozygous *met1-3* mutations led to a 32% decrease in CO frequency in *Arabidopsis* FTL CEN3. However, this effect was described as stochastic as the CO rate in the subsequent generation proved to be much more variable then in the original F_1 plants. The mild effect observed on the plants in this study is likely due to the impact of the chemical not being as strong as the impact of the homozygous mutations observed in *met1-3* mutant plants. These plants are characterised by high-sterility levels and severely impaired development, whereas the zebularine-exposed plants recovered when transferred to soil without chemical treatment.

Finally, recombination rates in the CTL3.9 interval were 14.5% lower in *Arabidopsis* F_2 seeds after zebularine treatment (80 µM) compared to the untreated control seeds (Figure 4.3.7), even though this difference does not appear to be significant (p = 0.0546).

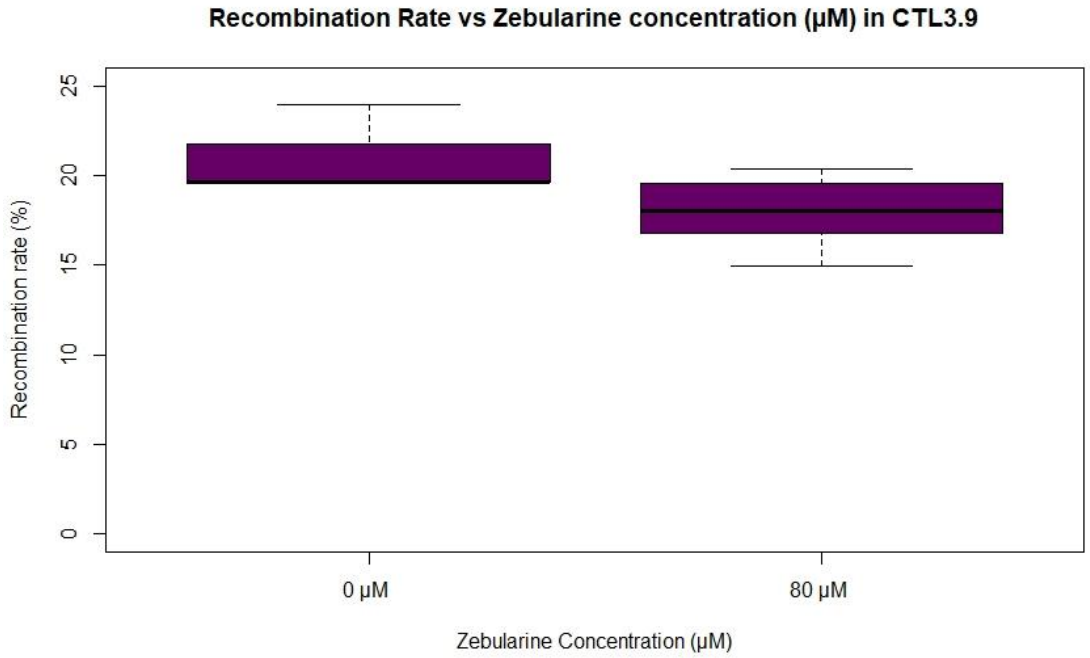


Figure 4.3.7. Zebularine-treated plants show a lower number of CO events in centromeric locus 3.9 (ANOVA, p = 0.0546) compared to untreated plants (n = 3 control, 8 treated, ≤500 seeds per plant)

As was discussed in 4.2.9, the quality of the fluorescence for the CTL3.9 seeds was too poor for the seed sets to be analysed automatically by CellProfiler. The contrast between fluorescent and non-fluorescent seeds was not strong enough for the green fluorescent marker, and impaired by high levels of auto-fluorescence. This means the data obtained for CTL3.9 was manually counted which accounts for fewer seeds analysed than for CTL4/20 and CTL2.2. The data should therefore be interpreted with caution as there is a potential for user-bias and error, and the statistical power of this data is also limited. However, it appears that the application of zebularine on CTL3.9 F_1 seeds leads to a reduction in CO frequency in the centromeric region albeit non-significantly. This is unexpected when comparing this result to what had previously been observed in *met1-3* mutants (Yelina *et al.* 2012), which showed an increase recombination rate in centromeric regions in the demethylated plants.

The discrepancies between the observed results in zebularine-treated plants and the described effects in DNA-methylation mutants *met1-3* and *ddm1* on recombination frequency in sub-telomeric regions and centromeric regions could be explained by the

fact that despite the assumption of the phenocopying of *met1* mutants, seedlings exposed to zebularine do not react in the same manner to the chemical treatment as they do to the mutation. It has previously been shown that *Arabidopsis* plants recovered near-normal levels of DNA-methylation 8 weeks after transfer to a chemical-free growth medium when initially treated with 40 µM zebularine (Baubec *et al.* 2009). However, the speed and the mechanisms by which the plants recovered was not clarified to this date. It is possible the plants would respond to being transferred to their recovery medium by hyper-methylating their genome during the meiotic phase, leading to the opposite effects and as a consequence globally decreased recombination rates in zebularine-treated F1 seedlings.

Zebularine does not only impact CG-methylation in *Arabidopsis* plants, but also has a decreasing effect on CHG and CHH (where H is A, T, G) methylation (Griffin *et al.* 2016). The genes *met1* and *ddm1* are mainly involved in the maintenance of CG-methylation (Zhang *et al.* 2018), and it is not clear yet to which extent the RNA-directed DNA-methylation (RdDM) pathway is impacting meiotic recombination. Male meiocytes are however characterised by a specific methylome, with higher levels of CG methylation but lower levels of CHG and CHH methylation (Walker *et al.* 2018). This led to the identification of Sexual-Lineage-specific methylated loci, which are differentially methylated in somatic cells and meiotic cells (meiocytes, microspores and sperm cells). It is therefore possible that zebularine interferes with genes involved in the RdDM pathway in addition to *met1* and *ddm1*, causing further disruption to the meiotic process and leading to the observed global diminution in crossover events in the genome.

These observations confirm that DNA-methylation plays a complex role on the regulation of meiotic crossover frequency in *Arabidopsis*, even though the extent and mechanisms of such an interaction are yet to be fully quantified and characterised. The rest of this chapter will focus on resolving whether similar effects can be observed in large genome cereals such as wheat and barley, as chemically manipulating the recombination landscapes in these species would be a major asset in traditional breeding and gene mapping for academic research.

4.3.6. The application of zebularine transiently impedes barley seedlings germination.

Similar to the work on *Arabidopsis* seedlings, barley F_1 hybrid seeds "Jettoo" were germinated on increasing concentrations of zebularine for 14 days as detailed in 4.2.14. Barley seedlings responded to the zebularine treatment in a dose-dependent manner. Seeds treated with high concentrations of zebularine showed a decreased germination rate compared to control seeds, and the plants exhibited a delayed development when treated with zebularine (Figure 4.3.8). Seedlings treated with 100 µM and 200 µM zebularine also exhibited shrivelled, curled leaves, compared to the untreated seedlings.

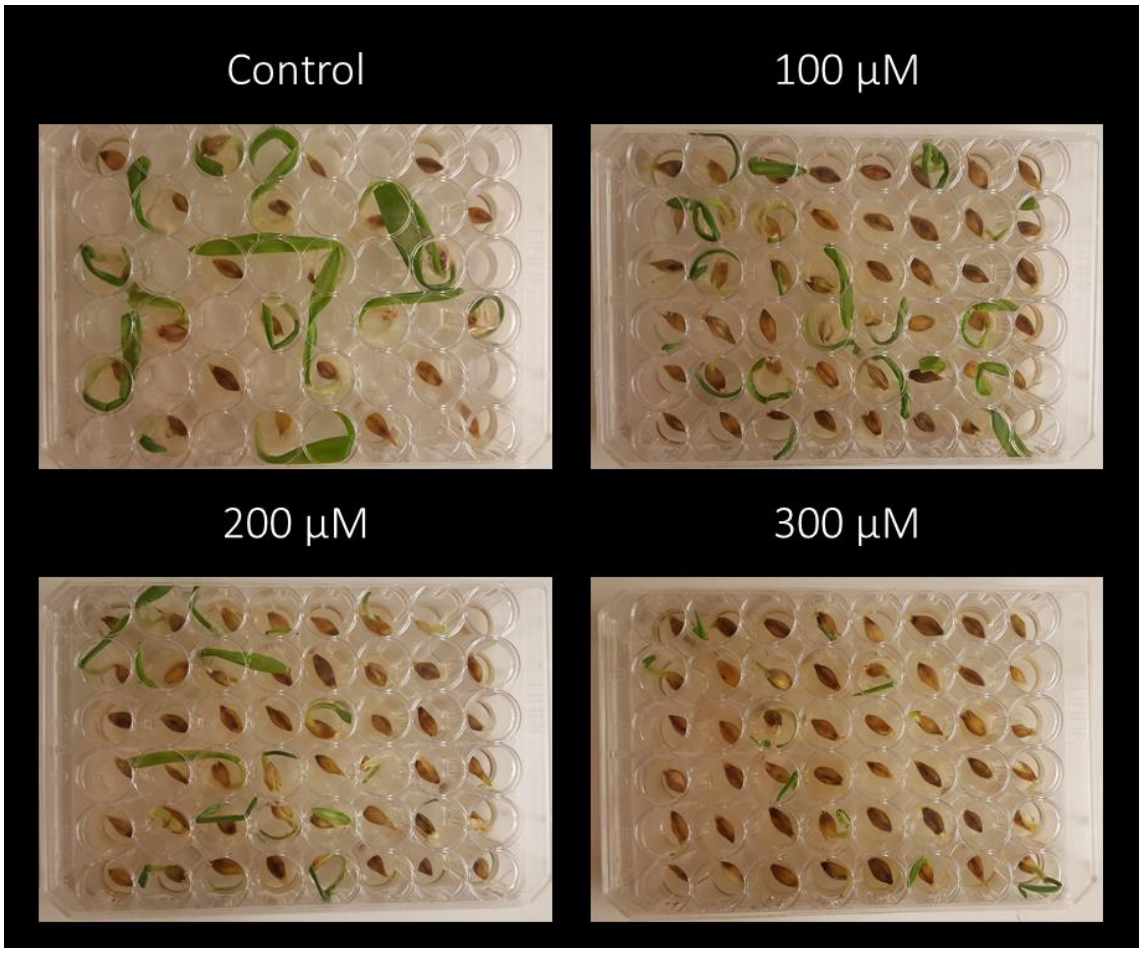


Figure 4.3.8. Dose-dependent response from barley F_1 hybrid (*cv.* Jettoo) seedlings to the application of increasing concentrations of zebularine. Germination rates are decreased and plant development delayed with higher concentrations of zebularine.

As described in 4.2.15, the seedlings germinated on zebularine medium were then transferred onto soil and grown in a controlled glasshouse environment. Jettoo being a winter F_1 hybrid, the seeds required a vernalisation period in a controlled growth cabinet (4°C, 16 hours day, 8 hours night) for 8 weeks, 2 weeks after being transferred to soil. The plants were then placed again in a controlled glasshouse environment for the remainder

of their development. As can be seen on Figure 4.3.9, the plants quickly recovered to normal phenotypes and no difference could be observed between plants which had been germinated on high concentrations of zebularine and the control untreated plants. All plants produced a normal seed set and no sterility could be observed, allowing the recovery of F_2 seeds for genotyping analysis.

Figure 4.3.9. Zebularine treated plants 16 weeks after transfer to soil. These plants were vernalised for 8 weeks, 2 weeks after transfer. The plants recovered a normal phenotype and produced normal seed sets.

The fact that these plants returned to a WT phenotype so early in their development was unexpected, as the zebularine concentrations used were high enough to cause very severe phenotypes. Given the semi-replicative nature of DNA-methylation, it was expected that part of the demethylated methylome would have been maintained as such even after the treatment was stopped and the plants would have exhibited similar phenotypes to those observed in *met1* rice plants such as high sterility, delayed development and heavy tillering (Hu *et al.* 2014). However, in tomato plants a single

mutation of DDM1 genes *Slddm1a* or *Slddm1b* was previously observed to not trigger a phenotype distinguishable from wild type plants (Corem *et al.* 2018). Only in double mutants *Slddm1a-Slddm1b* did the plants exhibit smaller leaves and underdeveloped fruits with little to no seed set. This shows the role of DNA-methylation in crops such as rice, tomato, and barley, varies between species and is not yet fully understood.

The Jettoo F_1 winter hybrid seedlings needed to be submitted to an 8-week vernalisation period to develop appropriately. This means the period between when the treatment was removed and when the plants developed seed was prolonged by 8 weeks. It is possible that zebularine degraded during this period and the levels of DNA-methylation were little to not affected during seed production. Additionally, being an F_1 hybrid developed by Syngenta, the Jettoo variety benefits from a strong hybrid vigour. The plants were very tall and produced large seeds in large quantities, compared to inbred malting spring cultivars commonly used in research such as Golden Promise and Barke. This might have helped the plants resist the effects of zebularine more than homozygous lines would have been able to, leading to no visible phenotypes after recovery. Finally, the concentration of zebularine might not have been sufficient to trigger large scale effects on the plants development and the seedlings might have needed to be exposed for longer to the chemical, or to higher concentrations post-germination, before being transferred into soil. Attempts were made to treat plants by injections later in their development to target meiosis. However, this technique proved difficult in practice in the greenhouse and would have required more optimisation. In the interest of time, this technique was not pursued, and focus moved to zebularine-treated seedlings and *met1* mutants generated by TILLING (See Chapter 5).

4.3.7. The application of zebularine to barley seedlings mildly impacts gene expression.

Gene expression analysis was performed to compare zebularine treated seedlings (300 µM) against untreated seedlings as described in 4.2.17. After the microarray run, a PCA plot (Figure 4.3.10) was obtained which showed a large level of variability between bio-replicates from the same conditions (treated and untreated plants).

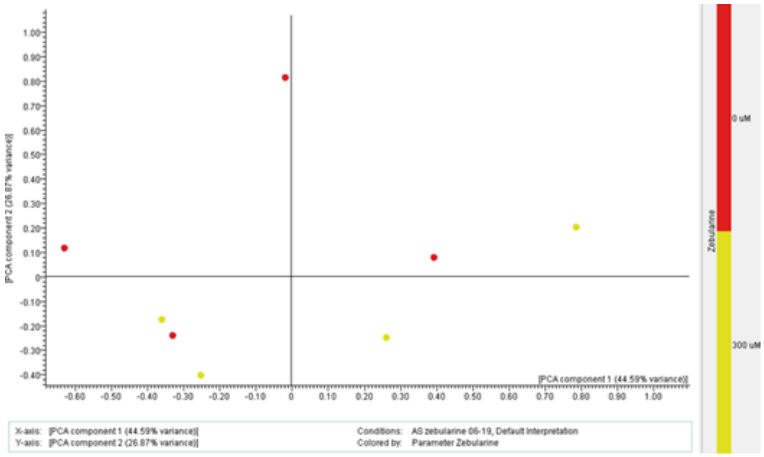


Figure 4.3.10. Principal Component Analysis (PCA) plot for the microarray gene expression analysis of plants germinated on 300 µM zebularine and untreated seedlings. The samples for both conditions are very scattered suggesting a large variability between bio-replicates.

The variability seen here is likely due to each bio-replicates consisting of a single plant and not a pool of different seedlings as had been done for the gene expression analysis of *Arabidopsis* zebularine-treated seedlings. Moreover, the seedlings were germinated on plates with zebularine and DMSO (300 µM zebularine-treated plants) or DMSO only. The application of zebularine on *Arabidopsis* plants was shown to trigger stochastic effects on plant development and gene expression (Baubec *et al.* 2014) and it is possible the application of DMSO also led to differential effects in untreated samples. It would have been ideal to repeat the experiment using pools of plants instead of singe individuals for each bio-replicates.

Fold-change analysis of gene expression between samples treated with 300 µM zebularine and untreated seedlings did nevertheless identify 309 genes that were differentially expressed in zebularine-treated plants ($p \leq 0.05$), with 40 of each having a *p*-value $p \leq 0.01$ (Supplementary Table 1). None of the genes traditionally involved in DNA-methylation mechanisms were identified as differentially regulated in zebularine-treated plants in either of the lists. However, a few genes of interest were identified in the genes with *p*-values under 0.05.

Lysine-specific demethylase 7A (MLOC_43699.1) is a protein involved in the demethylation of methylated Lysines in histone proteins, notably H3K9me2. In contrast, Histone-lysine N-methyltransferase, H3 lysine-9 specific 5, SUVH5 (MLOC_32627.1), is responsible for the methylation of Lysine 9 in H3K9me2. Surprisingly, both genes coding for these proteins were upregulated in zebularine-treated plants compared to mock-treated seedlings. As mentioned in Chapter 1, the H3K9me2 histone performs a crucial

role in the catalysis of CHG methylation as it binds to the Chromomethylase 3 (CMT3) protein through its Bromo-adjacent homology (BAH) Domain (Du *et al.* 2012). This interaction is paramount for correct binding of CMT3 to the nucleosome and methylation of the genome. In zebularine-treated plantlets, the regulation of H3K9me2 methylation is heightened, which could lead to a tightening of the CHG methylation regulatory process in demethylated plants. Another putative SUVH gene (MLOC_61815.1) is also upregulated in zebularine-treated seedlings, however its role in DNA-methylation has not been confirmed.

DSB repair constitutes a crucial part of the meiotic process, especially for the resolution of chiasmata into CO events. Genes coding for Double-strand-break repair protein rad21-like protein 1 RAD21l1 (MLOC_74114.2) and RAD51 analogue DNA repair protein radA-like protein RAD-A (MLOC_9144.1) were both upregulated in plants exposed to the zebularine treatment compared to untreated plants. These genes are involved in the repair of DSBs during meiosis. This could lead to changes in the patterns of recombination events in zebularine-treated plants with more chiasmata being repaired as CO events.

Finally, two other genes were identified as differentially expressed for which orthologues from the same family in *Arabidopsis* were also differentially regulated in seedlings treated with 5-azacytidine and zebularine (see 3.3.2 and 4.3.4). Germin-like protein 2a (MLOC_9957.3) is the barley orthologue of Germin-like protein 2 (GER2, AT5G39190) in *Arabidopsis*. GER2 is the same family of protein as GER3 (AT5G20630), for which gene expression was highly downregulated in zebularine-treated *Arabidopsis* seedlings as described in 4.3.4. Similarly, this gene is downregulated in barley plantlets exposed to zebularine compared to mock-treated plants. On the other hand, Flowering promoting factor-like 1 (AK251879.1) is an orthologue to the *Arabidopsis* Flowering-promoting factor 1 FPF1 (AT5G24860). FPF1 belongs to the same gene family as FLOWERING WAGENINGEN (FWA, AT4G25530) which is highly upregulated in *Arabidopsis* seedlings treated with 100 µM 5-azacytidine (Griffin *et al.* 2016). This gene was also upregulated in zebularine-treated barley seedlings compared to untreated samples.

Despite the large variability between bio-replicates, both treated and untreated, the highlighting of genes which were previously identified as differentially regulated in demethylated *Arabidopsis* plants supports that the tendencies observed through this microarray analysis are biologically consistent with what may be expected to be the effect

of DNA-methylation on gene expression in barley plants. Of course, these observations should be treated with a degree of caution, as the high level of variability between bio-replicates means no False Discovery Rate (FDR) correction could be applied in the analysis. This means any tendencies observed in this list should be further validated by RT-qPCR, similar to what was performed on zebularine-treated *Arabidopsis* plants (4.3.4).

4.3.8. Zebularine-treated plants are not characterised by a significant change in recombination.

After treatment of Jettoo F_1 hybrid seeds with increasing concentrations of zebularine, F_2 seeds were recovered and used for a recombination rate analyses using Single Nucleotide Polymorphism (SNP) markers. First, the original F_1 seeds were genotyped using a 50k SNP array in order to identify SNPs that were heterozygous in the Jettoo variety (4.2.18). A total of 10,048 markers were identified as heterozygous in Jettoo, which represents 22.9% of the total number of markers which are present on the 50k SNP chip. Chromosome 6H contained the most heterozygous markers for Jettoo (1,608 in total) and markers were more evenly distributed across the length of the chromosome. 24 SNP markers were therefore chosen on chromosome 6H, distributed informatively across the genetic map, which were used to perform the genotyping analysis on the F_2 populations (Figure 4.3.11).

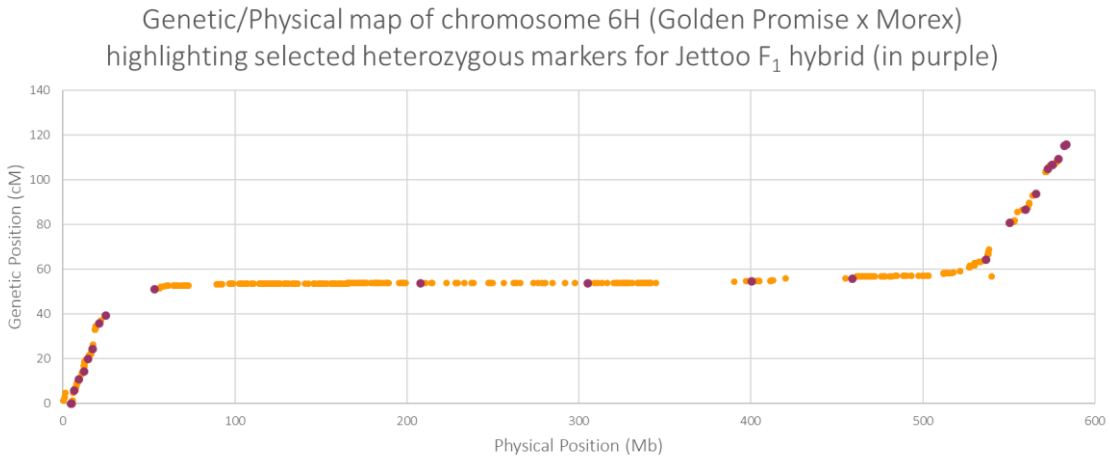


Figure 4.3.11. Position of heterozygous SNP markers on Chromosome 6H for Jettoo F_1 hybrid.

The genotyping analysis of the F_2 generation derived from treated and untreated Jettoo F_1 hybrids was carried out as described in 4.2.19. The overall number of CO events was not significantly impacted by the zebularine treatment. F_2 seedlings from untreated F_1 plants were characterised by an average number of COs of 2.54 on chromosome 6H, whereas the F_2 plants from treated F_1 seedlings had on average 2.56 CO events on chromosome 6H (p = 0.455). The recombination landscape was also unaffected by the zebularine treatment (Figure 4.3.12) when looking at the number of CO events in between each studied SNP marker across the chromosome. This means no displacement of CO hotspots was observed in zebularine-treated plants.

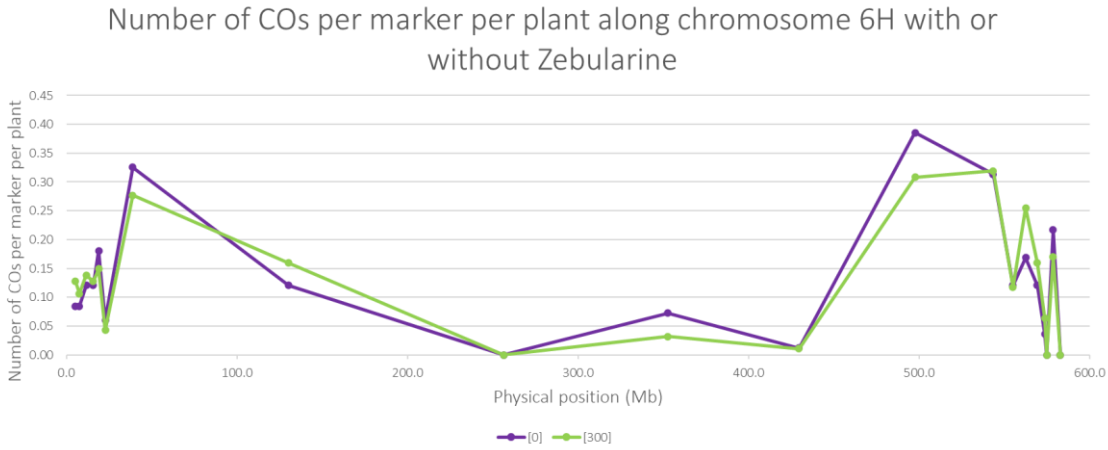


Figure 4.3.12. Distribution of CO events across Chromosome 6H in zebularine treated plants (300 µM, in green) compared to untreated plants (0 µM, in purple). No significant difference was evident in CO rates between treated (n = 94) and untreated (n = 83) barley plants.

As mentioned in Chapter 1, the recombination landscape in large genome crops is characterised by a reverse-bell shape with little to no recombination in the centromeric regions and the extremities of the telomeres (Baker *et al.* 2014, Saintenac *et al.* 2011). In contrast, recombination rates are higher in sub-telomeric regions. This is further confirmed by the data in Figure 4.3.12, where despite zebularine treatment, recombination rates were much higher between markers situated in the sub-telomeric region of Chromosome 6 and very low in the centromeric region.

The absence of any difference between treated and untreated F_1 plants could be explained by the plants having recovered a normal phenotype by the time the meiotic process was happening, as was described in 4.3.6. As mentioned, this could have been due to zebularine being degraded at the time of meiotic recombination, or the concentrations not having been high enough to have a significant impact on the meiotic machinery. It is therefore possible milder effects would not be identified and would

require a larger number of plants. Ideally, a larger number of F_2 plants would have needed to be analysed to identify smaller changes in recombination rates, and the 50k SNP chip should have been used to precisely identify where in the genome these effects were observed. However the use of this resource on such a large population was cost prohibitive.

4.4. Conclusion

In this chapter, zebularine, a demethylating agent, was used to alter the methylome in *Arabidopsis* and barley (*Hordeum vulgare*) to assess the effect of DNA-methylation on meiotic recombination patterns.

For both plant species, the application of zebularine triggered dose-dependent transient effects on development, leading to unhealthy-looking smaller plants which recovered to normal phenotypes after interruption of the treatment. In line with what had previously been described in the literature (Baubec *et al.* 2014, Griffin *et al.* 2016), zebularine-treated plants also exhibited changes in gene expression patterns in *Arabidopsis* and to a lesser extent, in barley. The milder effect observed in barley for the zebularine treatment on gene expression could have been caused by the large variability between bio-replicates from the same condition (treated and untreated seedlings). It is possible that a more consistent effect would be observed if pools of plants had been analysed for each bio-replicate instead of single plants in order to balance the stochastic effect that the germination treatment could have caused on the seedlings. Moreover, it would have been interesting to validate the trends identified in the microarray dataset by performing RT-qPCR on young barley seedlings germinated on similar concentrations of zebularine for the genes highlighted in 4.3.7 such as Lysine-specific demethylase 7A and SUVH5.

Surprisingly, zebularine-treated *Arabidopsis* FTL lines were all characterised with a tendency to have fewer CO events than untreated plants in all studied intervals. This was unexpected as previous descriptions of *met1* and *ddm1 Arabidopsis* mutant lines were characterised with higher CO rates in sub-telomeric and centromeric intervals (Melamed-Bessudo & Levy 2012, Yelina *et al.* 2012). As explained in 4.3.5, despite zebularine-treated *Arabidopsis* plants phenocopying *met1-3* mutants during vegetative growth, the DNA-methylation mechanisms impacted by zebularine do not solely concern the activity of the MET1 protein, but also affects the RdDM pathway (Griffin *et al.* 2016), which could lead to differential outcomes on the meiotic recombination landscape in these plants. The

observed tendencies would also need to be validated in a larger number of plants, for example using large-scale genotyping in F_2 lines after zebularine treatment in order to fine-tune where the CO events are happening in the genome.

No effect was observed on the recombination rates of zebularine-treated barley F_1 hybrid plants, compared to untreated F_1 plants. It is not clear whether the absence of effect is due to: (1) changes in the recombination landscape that are too mild to be observed on a low number of plants with only a few markers; (2) the long generation time of F_1 winter hybrid Jettoo leading to zebularine being degraded, and normal methylation reinstated, by the time meiosis is initiated in the spikes; (3) if DNA-methylation does not regulate meiotic recombination in barley in the same manner it does in *Arabidopsis*.

In this chapter, two main limitations were discovered in the study of the application of zebularine on DNA-methylation levels and meiotic recombination rates in plants. Firstly, as was described in 4.3.3 and 3.3.1, the DNA-methylation assay, which was developed during this PhD project, proved not to be robust enough to detect changes in DNA-methylation levels in zebularine-treated *Arabidopsis* and barley plants. This means the extent to which zebularine would have impacted the methylome in these plants has not been measured, and there is no certainty that DNA-methylation levels were impacted enough during the meiotic stages to have a significant impact on recombination levels. It would be interesting in the future either to use a different DNA-methylation assay technique, such as Whole Genome Bisulfite Sequencing (WGBS) or Methylated DNA immunoprecipitation (MeDIP), described in 3.1.1, or to further optimise the existing enzyme-based assay developed during this PhD project.

As the zebularine-treated barley plants returned to a normal phenotype when transferred into soil without zebularine it suggests that the effect of the treatment might have faded by the time the plants reached the initiation of meiosis. Several solutions to this problem were raised when developing the experimental plan on the best strategy to treat plants with zebularine, which involved incorporating zebularine to the watering system, spraying the plants with the chemical through their developmental process, or directly injecting the plants with a zebularine solution pre-meiotically. However, these methods have both cost and safety implications. Trials of injecting plants with zebularine in the stem prior to meiosis were made, however it was difficult to assess whether zebularine was injected at the right stage and indeed penetrated the spikes without

invasively dissecting the flower out of the stem. This method would have proved to be too labour-intensive and required optimisation, therefore it was decided not to proceed and focus instead on generating stable *met1* barley mutant plants using a large TILLING population in order to characterise the role of DNA-methylation on meiotic recombination.

5. Natural and induced variation in met1 and ddm1 in Barley

5.1. Introduction

Creating genetic diversity in plants is a key strategy for improving the agricultural industry in the future years. As discussed in Chapter 1, there is an ever-growing demand for developing crops with enhanced yields, pests and abiotic stress resilience and reduced need for agronomical inputs. As the yields of our broad-acre crops have effectively plateaued, this presents a complex challenge for agricultural industries, especially crop breeders, as they have now reached a bottleneck using traditional resources and need new strategies to generate new varieties.

Discovering and characterising variation within genes has been a foundation of discovery research. Genetic variation in fundamental research is used in different ways: by noticing an unusual phenotype in a population and using genetic approaches to map the causal gene(s), or by identifying a gene of predicted potential interest in the genome and knocking it out or down in order to understand its role and how it operates at the molecular level. By mutating genes, their function and place in biological mechanisms and processes can be better understood.

Whether for commercial purposes or academic research, discovering or creating lines containing novel variants of genes is crucial and relies on different mutagenesis techniques. These can be divided in two categories, depending on the method used and the end-product. Targeted mutagenesis technics can use transgenic technologies such as CRISPR-Cas9 gene editing or the delivery of hairpins to specifically target and silence genes of interest. Plant transformation is also widely performed to overexpress genes of interest in order to characterise their effect either in wild-type plants or mutated plants which carry a non-functional allele. On the other hand, random physical or chemical mutagenesis relies on random disturbances of the genome by a range of mutagens (UV-radiations, γ-radiations, ethyl methanesulfonate (EMS), N-ethyl-N-nitrosourea (ENU)). These techniques are commonly used in reverse genetics to select for a phenotype of interest and investigate the genes causing it. In parallel, there is a growing interest in the

use of ancestral varieties and the study of natural variants for a trait of interest. These represent a large pool of genetic diversity, for example to re-introduce resistance genes in modern cultivars. In general, mutations caused by targeted mutagenesis would lead to complete knockouts of the gene, causing severe phenotypes. There is however a growing interest in using gene editing techniques to fine-tune edits in genes of interest at a single nucleotide level, for example using Homology Directed Recombination (HDR) associated to the CRISPR-Cas9 technology. Mutations caused by random mutagenesis techniques will often trigger milder effects on protein function as they will often consist of Single Nucleotide Polymorphisms, however a large number of additional mutations will also be present in the mutant plant. There is therefore a need for several backcrosses with the wild-type cultivar in order to eliminate these additional mutations, making the use of such mutants time-consuming. Finally, natural variants contain specific alleles, which usually cause a mild effect on the genes of interest which enables the plants to grow in the wild without severe impacts on their phenotypes.

5.1.1. Targeted mutagenesis techniques in Barley

When a gene has been widely characterised and identified as a good candidate for controlling a given phenotype, the best way to validate its function or to identify its role in a pathway is frequently to knock the gene out completely. Causing the gene to be non-functional in a plant may often cause a severe and easily traceable phenotype using modern genetics. Attempts to restore the wild-type phenotype by re-introducing a functional version of the gene or its homologue will then confirm the gene's role in the studied biological system. More detailed interactions, for example with other components in a molecular pathway, can be investigated by altering defined protein domains. This can help identify whether and how a protein interacts with other proteins. One of the best options to manipulate specific regions of a gene is to use genetic modification techniques, which allow specific targeting of defined locations within target genes. Genetic modification can similarly be used to validate a mutant gene function by complementation, by introducing its wild-type version, or an orthologous version from another species. For example, recently, complementation of sterile meiotic *swi1* mutants with domain swapped versions of *Swi1* with maize and rice orthologues restored pollen viability and fertility in *Arabidopsis* plants and highlighted the conservation of this protein's structure and function across plant species (Mahesh *et al.* 2021). However, in barley, due to most cultivars being incompatible with the current genetic modification

techniques, complementation assays on induced or natural variants are generally not viable (Schreiber *et al.* 2019).

Several different targeted mutagenesis techniques are currently being used in barley including virus induced gene silencing (VIGs), RNA-interference (RNAi), TALENs and CRISPR-Cas9. All are based on transgenic biology – generally achieved through Agrobacterium mediated approaches. In addition, the use of reporter lines and tissue-specific promotors provide information on gene expression patterns at the cellular level and the ability to knock out/down genes in specific tissues at specific times.

5.1.1.1. Virus-induced gene silencing (VIGS)

VIGS is a method ideally used for non-model species for which transformation is limited or even impossible. VIGS allows the study of a gene's function by silencing it in a transient way. This means VIGS is an ideal alternative for knocking a gene down when the outcome of stable transgenics may be predicted to be lethal or lead to complete sterility. Briefly, a fragment of the gene of interest is introduced into a viral vector and transformed in *Agrobacterium tumefaciens*. *Agrobacterium* is then infiltrated in the tissue selected for the study, usually leaves, and the gene of interest will be targeted by the endogenous plant RNA-induced Silencing Complex (RISC) for silencing via degradation of the mRNAs, leading to a decrease in mRNA available for protein translation. In barley, Barley Stripe Mosaic Virus (BSMV) is the main vector used, but it can also be used to transform other species such as wheat, rice, or maize. For example, a BSMV VIGS construct was successfully used to study the effect of Brassinosteroid Insensitive 1 (BRI1) on leaf resistance to Fusarium head blight (FHB) (Gunupuru *et al.* 2019). In this study, VIGS silencing of *BRI1* effectively rendered plants (*cv*. Akashinriki) more susceptible to *Fusarium culmorum* infection, which translated into larger areas of infection on leaves infiltrated with the BSMC VIGS construct targeting *BRI1* in comparison to mock and empty vector controls.

5.1.1.2. RNA-interference (RNAi)

Another method for silencing a gene is the use of RNAi lines. These lines are genetically modified to express a short hairpin RNA (siRNA) which will bind to the mRNA of the gene interest with the help of an RNA-induced Silencing Complex (RISC), similarly to the VIGS technique. This will lead to cleavage of the mRNA and absence of translation into a functional protein (Agrawal *et al.* 2003). RNAi lines were used in 2018 to investigate the

role of Caffeic acid O-methyltransferase (COMT), a lignin biosynthesis gene of major importance for digestibility of the straw (Daly *et al.* 2019). The silencing of mRNA in these lines led to a change in the structure of lignin which makes them particularly interesting for use in animal feed and biofuels production.

5.1.1.3. Transcription activator-like effector nucleases

TALENs consist of the association of Transcription activator-like (TAL) effectors with the non-specific endonuclease *FokI*. There is a large variety of TALs specific to certain pairs of nucleotides, which means when combining different TALs the resulting complex will bind specifically to unique DNA sequences. When coupled to *FokI*, a TALEN complex targets specific sequences in the gene of interest and creates DSBs. DSBs will then be repaired by the notoriously error prone non-homologous end joining (NHEJ) mechanism. This will likely lead to mutations in the gene of interest. Despite its high specificity, the TALEN system relies on protein-DNA interactions, making it a long and expensive process, especially in large genome crops like barley. For this reason, only a limited number of studies have used the TALEN system to genetically engineer genes of interest. In 2017, TALEN constructs were used to generate mutations in the promotor of a barley phytase gene of the purple acid phosphatase group (*HvPAPhy_a*) (Holme *et al.* 2017). This led to a reduction in *HvPAPhy_a* expression and validated this gene as one of the main actors in Mature Grain Phytase Activity (MPGA), a crucial process involved in seed germination.

5.1.1.4. CRISPR-Cas9

Finally, the favoured method for genetic transformation is the use of a CRISPR-Cas9 construct, which allows direct targeting of where the mutation will occur. The CRISPR-Cas9 complex consists in a single guide RNA (sgRNA) fused to a Cas9 nuclease. The sgRNA, complementary to a specific sequence in the target gene, will lead the nuclease to the gene of interest, which will lead to the production of DSBs in the gene of interest. As with the TALEN system, the DSBs will be repaired by non-homologous end joining mechanisms, known to be extremely error-prone, which will result in mutations introduced at the DSB formation site. One of the main advantages of this method is it allows the creation of several mutant lines with distinct mutations using the same construct whilst mitigating the effect of the mutation on the protein's function. Additionally, the Cas9 gene can easily be eliminated from the lines by genetic segregation, making them almost identical to mutant lines generated with non-invasive methods.

Therefore, only the effect of the mutation will cause the novel phenotypes. Compared to the TALEN method, CRISPR relies on RNA/DNA interactions that are more specific and stronger than protein/DNA interactions. The method is faster, easier to perform and more cost efficient than the use of TALENs. Many studies have used CRISPR to generate mutant lines in barley. Recently, CRISPR-Cas9 was successfully used to generate a range of mutants in (1,3;1,4)-β-glucan synthase genes HvCslH1, HvCslF3, HvCslF6 and HvCslF9 (Garcia - Gimenez *et al.* 2020). *cslf6* mutants were characterised with an absence of (1,3;1,4)-β-glucan in the grains, in association with alterations to Thousand Grain Weight (TGW) and grain morphological traits. Similarly, CRISPR mutants in Apetala2 gene HvAP2 resulted in a distinctive *gigas1.a* phenotype on grain development leading to altered floret organ identity and abnormal growth of lemmas and paleas (Shoesmith *et al.* 2021).

5.1.2. Creating genetic diversity by random mutagenesis

As discussed in Chapter 1, the use of genetically modified lines in commercial breeding is very limited. Globally, no transgenic barley lines are currently being grown for commercial purposes. Breeders must rely on other methods to generate genetic diversity and create new varieties. One method for achieving this is through random mutagenesis techniques such as radiation or chemical treatments. These approaches will induce many mutations randomly in the genome in a dose-dependent manner. In addition to their appropriateness for the crop breeding industry, random mutagenesis techniques also provide large advantages for fundamental research, due to causation of novel and interesting phenotypes. This is also an opportunity to create a large databank of mutant lines rapidly, which can then be studied depending on the researcher's interests.

Random mutagenesis in plants was first described in 1928 when L. J. Stadler published a study showing that ionizing radiation could induce mutations in maize (Stadler 1928). This opened the door to the generation of many more mutant lines using various radiation techniques including X-rays, UV-rays, and Gamma-rays. Golden Promise, which was one of the most cultivated barley cultivars in the late 20th century in Scotland for its malting properties, was obtained in 1956 by gamma-ray radiation of the salt-tolerant cultivar Maythorpe (Forster 2001). Currently, Golden Promise is still widely used in academic research as it represents one of the only cultivars susceptible to *Agrobacterium*-mediated genetic transformation (Hisano & Sato 2016). X-ray mutagenesis led to the production of *cv.* Diamant in 1965 (Bouma 1967). This cultivar was characterised by a 12% yield increase

compared to its parent Valticky and was used to produce more than 150 hybrids, making its economic and scientific impacts virtually incalculable.

Chemically-induced random mutagenesis started being used in the mid-1940's after researchers became aware of the mutagenic capabilities of some compounds and their advantage over radiation-induced mutagenesis. The application of chemicals on plants and seeds is more practical than the use of radiation and presents fewer risks for the operator. Many chemical compounds have been used for mutagenesis such as ethyleneimine (EI), diethyl sulphate (dES), EMS, N-ethyl-N-nitrosourea (ENU), N-nitroso-N-methylurea (NMU) and sodium azide (NaN_3). These mutagens mostly cause point mutations in the genome such as transitions and transversions, but occasionally other effects have been reported such as small insertions or deletions. EMS and NaN_3 have been widely used to create large mutant populations in barley, in various cultivars such as Optic (Caldwell *et al.* 2004) and Barke (Gottwald *et al.* 2009) for EMS and Lux (Lababidi *et al.* 2009) and Morex (Talamè *et al.* 2008) for NaN_3. These populations have been used for both forward and reverse genetics, notably using the Targeting Induced Local Lesions IN Genomes (TILLING) methodology to screen for mutations in genes of interest.

The primary limiting factor for mutants obtained by chemical and radiation mutagenesis is the presence of background mutations in addition to the mutation of interest. This means direct comparison of the newly discovered mutant allele with the wild-type cultivar is impossible and highlights the need for several rounds of backcrosses with the parenting cultivar to remove background mutations. Thus, several Nearly Isogenic Lines (NILs) were produced by recurrent backcrossing of several mutant lines of interest with the cultivar Bowman (Druka *et al.* 2011). This produced an invaluable resource of mutant lines which is still globally used for academic research and the identification of new genes of interest (Yang *et al.* 2021, Zhong *et al.* 2021, Poursarebani *et al.* 2020, Huang *et al.* 2020, Gol *et al.* 2021, Matyszczak *et al.* 2020, Colas *et al.* 2019).

5.1.3. Natural variants

Up until about 150 years ago, plant breeding was mostly done by farmers via domestication and selection of varieties best adapted to their environment and farming techniques (Bradshaw 2017). Modern-day breeding is based on the introgression of traits of interest into cultivated landraces from a wide range of other varieties, and relies increasingly on the help of molecular biology tools such as Marker-Assisted Selection

(MAS) and Next Generation Sequencing (NGS) to identify beneficial alleles in large panels of cultivars (Miedaner & Korzun 2012). These breeding techniques primarily depend on the natural diversity of alleles between varieties, making the conservation and analysis of ancestral and wild relatives of crops of paramount importance to breeding programs. In particular, in barley, a large collection of 256 wild barley accessions was analysed using molecular markers such as Single Nucleotide Polymorphisms (SNPs) and chloroplast simple sequence repeats (cpSSRs). This highlighted strong links between past climate change events, geographical positions of barley ecotypes, and their associated genetic variation levels (Russell *et al.* 2014). Similarly, the analysis of Genebank material from over 22,000 accessions allowed for the identification of a locus involved in awn roughness in barley on chromosome 5H, which was narrowed down to the identification of an orthologue to rice gene Long and Barbed Awn1 (LABA1) (Milner *et al.* 2019)

One example of successful traditional breeding is the identification of the Mildew resistance locus o (*Mlo*) in an Ethiopian wild barley variety, for which loss of function confers resistance to Powdery Mildew (Jørgensen 1992). This resistance is recessively-inherited and targets a broad spectrum of fungal strains of mildew, making it non-race specific (Lyngkjaer & Carver 2000). Several similar phenotypes were reproduced by random mutagenesis of the Mlo gene, with over 40 mlo mutants currently listed. Notably, natural mutant *mlo-11* (Piffanelli *et al.* 2004) and induced mutant *mlo-9* (Schwarzbach 1967) were extensively used in breeding programs, and it is now estimated that over 50% of the commercially available spring barley varieties are resistant to powdery mildew thanks to the introgression of these mlo alleles (Kusch & Panstruga 2017). Similarly, several mutant plants containing semi-dwarfing genes have been characterised and used in barley breeding program for their importance in lodging resistance, which also indirectly influences yield, grain quality and malting quality (Kuczyńska *et al.* 2013). The *sdw1* semi-dwarf allele, first identified in the barley cultivar Jotun (Foster 1987), is an allelic version of *denso*, which was obtained by EMS mutagenesis of cv. Abed Denso (Haahr 1976). Both alleles were extensively used to develop short-statured cultivars less susceptible to lodging, which were of particular interest to the malting industry (Kuczyńska *et al.* 2013). These two examples highlight the importance of the maintenance and characterisation of large germplasm collections, which represent invaluable pools of genetic diversity for future breeding programs.

5.1.4. Aims of the chapter.

Currently, no *met1* nor *ddm1* mutations have been characterised in barley. In this chapter, I aimed to use a novel TILLING by sequencing method to identify new mutants in *met1* amongst a large EMS-mutagenized population. These selected mutations were then isolated from the remaining mutations by crossing into a clean background (*cv.* Barke) before characterising the effect of the mutation on recombination. In parallel, a wide collection of wild barley and landraces was used to identify naturally occurring mutations, using previously published Exome capture data as a resource (Russell *et al.* 2014). Finally, the generation of CRISPR-Cas9 mutant lines for *met1* and *ddm1* was initiated in barley (*cv.* Golden Promise).

5.2. Material and Methods

5.2.1. *In silico* characterisation of methylation genes *met1* and *ddm1* in barley

Sequences for *Met1* and *Ddm1* were obtained from the newly assembled genome sequence for *cv*. Golden Promise (Schreiber *et al.* 2020) from Dr. Miriam Schreiber at the James Hutton Institute. In total, one copy of *Met1* was identified on chromosome 2H, but *Ddm1* was present in 3 copies on chromosomes 2H, 4H and 7H. Expression profiles for each gene copy were also obtained from RNAseq data available on the Barlex genome browser (Colmsee *et al.* 2015). CDS sequences were aligned on whole gene sequences to determine the intron/exon structure of the genes using the NCBI Splign online tool (Kapustin *et al.* 2008). Functional Domain Prediction was carried out using the InterPro Database online tool (Blum *et al.* 2021) and the identification of conserved amino-acids across plant species was performed on the Clustal Omega online database (Madeira *et al.* 2019).

5.2.2. Plant material

Before the start of this PhD project, the TILLING population had been generated by two rounds of EMS treatment. First, three lots of 20,000 seeds (*cv.* Golden Promise) had been mutagenized using 25, 30 and 35 mM EMS (M_0) as described by Caldwell *et al.*, (2004). The M_1 generation was grown in the field and bulk harvested. Then, 20,000 of the resulting M_2 seeds were treated again with 25 mM, 30 mM and 35 mM EMS and grown in the field to generate a pseudo M_1*. M_2* seeds were bulk harvested and grown in square 8 cm pots in 16x16 pot arrays for a total of 3,072 individual plants. At the 2-leaf stage, DNA was extracted by pooling plants in rows and columns for each block of 16x16 plants (256 individual plants). Three blocks were placed into one 96-well plate (768 individual plants per plate). Details of the pooling strategy are explained in Figure 5.2.1. DNA extraction was carried out using DNeasy Plant Maxi kit (Qiagen) and the resulting DNA was quantified by PicoGreen before being diluted to 5 ng/µL.

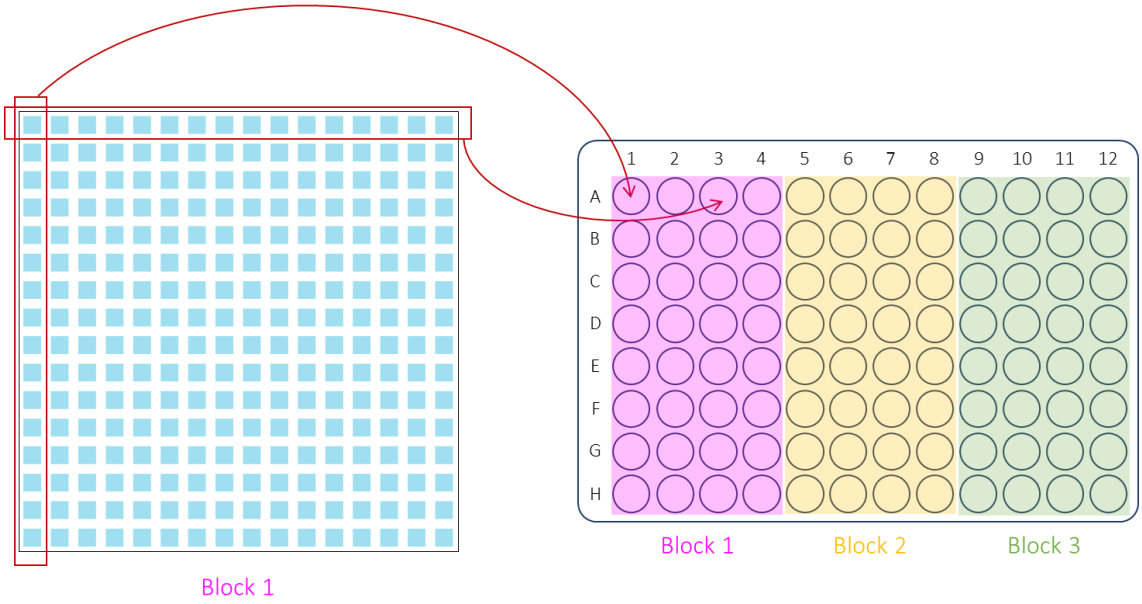


Figure 5.2.1. Plants were grown in 16x16 matrix in the greenhouse. DNA extractions were done in 96-well plates by pooling together plants from the same columns, then from the same rows. This resulted in each plate containing genetic material of 768 plants in total, representing 3 distinct matrices or blocks. In the example of the first DNA plate (Plate 1), it contains Blocks number 1, 2 and 3. Plate 2 will gather all 768 individuals from blocks 4, 5 and 6, and so on until Plate 4, totalling 3,072 plants.

5.2.3. Optimising library preparation

In total 14 pairs of primers were designed along *met1* (Figure 5.2.2). The primers were created to cover the conserved Bromo Adjacent Homology (BAH) and C-5 cytosine-specific DNA-Methyltransferase domains. In addition, the main intron-exon regions were

screened with the hope that mutations in these specific areas would lead to a complete knockout of the gene.

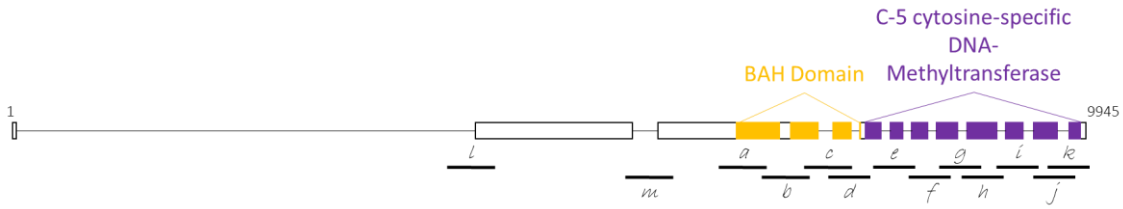


Figure 5.2.2. Structure of *met1* and position of the conserved BAH and methyltransferase domains. Each black line represents an amplicon for screening the gene for mutations. Letters from *a* to *m* match the primers listed in Appendix 8.1.

PCR conditions were tested to find the optimal annealing temperature for the designed primers using the KAPA HiFi HotStart ReadyMix PCR Kit (Roche) as described in 2.2.2.2. Multiplexing conditions were tested using KAPA DNA Polymerase and Q5 Hi-Fi Hot Start DNA Polymerase (New England Biolabs) following the protocols detailed in 2.2.2.1 and 2.2.2.2.

5.2.4. Multiplexing PCR with 13 amplicons

All primers resulting in the amplicons described in Figure 5.2.2 were pooled together from the original 100 mM stock solution (Sigma Aldrich) following the manufacturers' instructions. The final concentration of each primer was 3.84 mM in the primer mix. PCR reactions were carried out using Q5 Hi-Fi Hot Start DNA Polymerase (New England Biolabs) as follows: 5 µL Q5 Reaction Buffer (5X); 0.5 µL CleanAmp dNTPs (10 mM, Sigma Aldrich); 0.25 µL Q5 Hi-Fi Hot Start DNA Polymerase (2U/µL); 1.3 µL Primers Mix (3.84 mM, for a final concentration of 0.2 mM), 5 µL DNA template (5ng/µL, cv. Golden Promise) and sterile water to complete to a volume of 25 µL per reaction.

The samples were submitted to the following PCR program: Initial denaturation at 98°C for 30 seconds; 35 cycles of denaturation at 98°C for 10 seconds, annealing at 62°C for 15 seconds and extension at 72°C for 45 seconds; final extension at 72°C for 2 minutes; holding stage at 8°C.

The amplicons were then visualised on an agarose gel as described in 2.2.6.

5.2.5. Multiplexing PCR with non-overlapping amplicons

Two multiplexing mixes were carried out in order to separate the overlapping pairs of primers as described in the table of primers (Appendix 8.1) and on Figure 5.2.3.

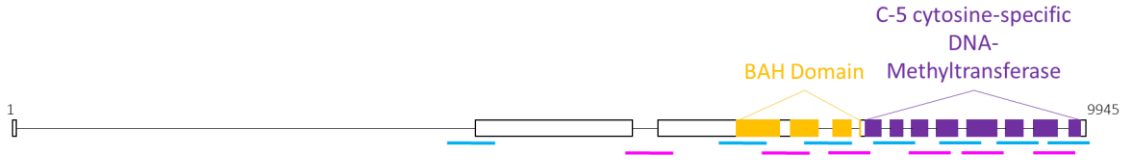


Figure 5.2.3. Two multiplex mixes were put together which allowed separation of the overlapping pairs of primers. Multiplexing mix 1 is depicted by the blue lines, and multiplexing mix 2 is represented by pink lines.

Multiplexing mix 1 (M_1) was composed of seven pairs or primers for a final concentration of 7.14 mM for each primer. Multiplexing mix 2 (M_2) contained six pairs of primers and was diluted with sterile water to achieve the same concentration of 7.14 mM for each primer. PCR reaction mixes included: 5 µL Q5 Reaction Buffer (5X); 0.5 µL CleanAmp dNTPs (10 mM, Sigma Aldrich); 0.25 µL Q5 Hi-Fi Hot Start DNA Polymerase (2 U/µL); 0.7 µL Primers Mix 1 or 2 (7.14 mM, for a final concentration of 0.2 mM), 5 µL DNA template (5ng/µL, cv. Golden Promise) and sterile water to complete to a volume of 25 µL per reaction.

Both PCR reactions with Multiplexing mixes 1 and 2 were completed in parallel as described in 5.2.4 and the amplicons were then visualised on an agarose gel (2.2.6).

5.2.6. PCR Clean-Up I

The first clean-up of the initial PCR reactions was carried out following the 16S Metagenomic Sequencing Library Preparation protocol supplied by Illumina but adapted to the needs of this experiment and to include the two multiplexing mixes. Briefly, 12.5 µL of each PCR Multiplexing mix 1 and 2 were added to 20 µL of AMPure XP magnetic Beads (Beckman Coulter) in a 96-well MIDI-plate. The samples were shaken at 1,400 rpm for 2 minutes before being incubated for 5 minutes at room temperature. The samples were then processed following the initial protocol from Illumina. Briefly, the MIDI-plate was placed on a magnetic plate stand (Invitrogen) and left to stand for 2 minutes until the AMPure XP magnetic beads had agglomerated against the magnet, leaving the supernatant clear. The supernatant was then removed carefully without disturbing the pelleted beads, and the samples were washed by adding 200 µL of freshly prepared 80% ethanol and leaving them to stand for 30 seconds. The supernatant was then carefully removed, and the wash step was repeated a second time. Excess ethanol was then thoroughly removed without disturbing the beads and the samples were left to air-dry for 10 minutes at room temperature. The MIDI-plate was then removed from the stand and the beads were resuspended in 52.5 µL of 10 mM Tris (pH 8.5). The plate was

then shaken at 1,400 rpm for 2 minutes and incubated for 2 minutes at room temperature. The samples were placed back on the magnetic stand and left to stand for 2 minutes until the supernatant was clear. 50 µL of supernatant containing the eluted DNA was then transferred to a fresh 96-well PCR plate, and correct cleaning of the samples was verified by electrophoresis gel migration (2.2.6).

5.2.7. Testing primer efficiency using qPCR.

The efficiency of each primer set was assessed using qPCR in order to check each amplicon was equally represented in the first PCR mix after amplification.

qPCR was carried out using Q5 DNA Polymerase (New England Biolabs) with the following protocol: 5 µL Q5 Reaction Buffer(5X); 0.5 µL dNTPs (10 mM, Promega); 0.25 µL Q5 DNA Polymerase (2 U/µL); 1.25 µL of Forward/Reverse Primers Mix (10 mM); 1 µL SYBR Green; 1 µL DNA template from PCR Clean-Up I (undiluted, 1/100 dilution or 1/1000 dilution); Sterile water up to 25 µL per reaction.

The samples were then processed for qPCR on a StepOne machine (Life Technologies): Initial denaturation at 98°C for 30 seconds with an initial fluorescence reading; 40 cycles of denaturation at 98°C for 10 seconds, annealing at 62°C for 45 seconds coupled to a fluorescence reading, extension at 72°C for 45 seconds; a final extension at 72°C for 2 minutes.

Results of the qPCR were analysed using the StepOne Software (v2.3, Life Technologies).

5.2.8. Indexing the PCR products

The indexing reaction consists of the addition of a barcode to the PCR products from the first amplification. This allows for data deconvolution after sequencing.

Indexing was performed following the 16S Metagenomic Sequencing Library Preparation protocol from Illumina using the Nextera Indexing kit A for 96 samples (Illumina).

In brief, 5 µL of cleaned up DNA (5.2.6) were transferred to a fresh 96-well PCR plate and were mixed with the following PCR reaction: 25 µL 2X KAPA HiFi HotStart ReadyMix, 5 µL Nextera XT Index Primer 1 (N7XX), 5 µL Nextera XT Index Primer 2 (N5XX), 10 µL sterile water. Each sample contained a unique combination of Primers 1 and 2, following the guidelines provided by the manufacturers.

The samples were then submitted to the indexing reaction using a thermocycler as follows: Initial denaturation at 95°C for 3 minutes; 8 cycles of denaturation at 95°C for 30

seconds, annealing at 55°C for 30 seconds, extension at 72°C for 30 seconds; a final extension at 72°C for 5 minutes.

The samples were then submitted to a second Clean-Up step as detailed in the Illumina protocol. Briefly, 50 µL of the indexing reaction samples were transferred to a 96-well MIDI-plate along with 56 µL of AMPure XP Beads. The plate was then shaken at 1,400 rpm for 2 minutes before being left to incubate at room temperature for 5 minutes. The plate was then placed onto a magnetic stand for 2 minutes until the supernatant was clear. With the plate still on the stand, the supernatant was discarded, and two washes were performed by adding 200 µL of freshly prepared 80% ethanol, leaving the samples to incubate for 30 seconds, and removing the supernatant. The leftover ethanol was then thoroughly removed, and the samples were air-dried for 10 minutes at room temperature. With the plate off the magnetic stand, 27.5 µL of 10 mM Tris (pH 8.5) were added to the samples and the beads were resuspended by shaking the plate for 2 minutes at 1,400 rpm. The samples were left to stand for 2 minutes at room temperature before transferring the plate back onto the magnetic stand. After 2 minutes, when the supernatant was clear, 25 µL of samples containing the eluted DNA were transferred to a clean 96-well PCR plate. A selection of samples was then migrated on an electrophoresis gel to verify correct amplification and cleaning of the indexed amplicons (2.2.6).

5.2.9. Library quantification, normalisation, and pooling

After indexing, the final clean amplicons were quantified using PicoGreen by the genomics facility in the James Hutton Institute. The samples were then diluted to 4 nM and pooled together in a single tube, 5 µL of each sample. The library was then given to the sequencing facility at the James Hutton Institute where its quality was assessed on a Bioanalyser before the library was processed on an Illumina MiSeq sequencer.

5.2.10. Analysing sequencing data

A custom reference was built using the *HvMet1A* gene sequence from the Golden Promise genome sequence assembly. Sequencing reads were first checked for consistency within the library to ensure the samples were all represented equally. The quality of the reads was then assessed using FastQC (version 0.11.8, http://www.bioinformatics.babraham.ac.uk/projects/fastqc) before removing the barcodes and poor quality bases with Trimmomatic (parameters: Leading 30; Trailing 30; Minlen 100; Bolger, Lohse, & Usadel, 2014). After trimming, the curated *met1* reference

sequence was used to map the reads using bwa mem (v0.7.10). Samtools (Danecek *et al.* 2021) was used to filter the alignments (cut-off of 220 for minimum alignment score), then sort them. Variant calling was completed with Freebayes (v.0.9.18) (Garrison & Marth, 2012) on each of the files separately with the following settings: haplotype length of 0 for annotating variants; minimum of alternate alleles called of 30; minimum fraction of alternate alleles called of 0.02; output all alleles which pass input filters; ignore complex events; ignore Multi Nucleotide Polymorphisms (MNPs); turn off left-alignment of indels; remove insertion/deletions events; assume samples result from pooled sequencing.

Due to the pooling strategy discussed in 5.2.2, Bcftools (Danecek *et al.* 2021) was used for an initial filtering of the variants, selecting for individual variants per column per block, or per row per block. Only SNPs which occurred within one block twice, once in a row and once in a column, were retained. Finally, filtering was applied to the variants to keep only the mutations which would be expected to be triggered by EMS treatment, G>A or C>T.

PROVEAN (Protein Variation Effect Analyzer, http://provean.jcvi.org/index.php) (Choi *et al.* 2012, Choi & Chan 2015) was used as a tool to determine whether the identified mutations would likely lead to a deleterious effect on the functionality of the protein. The program was run on a local server with the non-redundant NCBI database (updated October 2018). The Provean score is calculated by evaluating the effect the mutation has on the chemistry of the protein when causing an amino acid change, as well as how well conserved the concerned amino acid is amongst a large range of species. The lower the score is, the higher the probability that the mutation will have a deleterious effect on the protein's function.

5.2.11. Validating and following the mutations

For each plant identified as a candidate mutant by MiSeq or Exome Capture, eight progeny seeds were germinated in small pots and DNA was extracted by the genomics facility in the James Hutton Institute at the 2-leaves stage (2 weeks old). PCR was performed as described in 2.2.2.1 using the sequencing primers flanking the region where the mutation of interest has been identified. PCR products were purified either as described in 5.2.6 or 2.2.3 depending on the number of samples and sent to the genomics facility in the James Hutton Institute for Sanger sequencing. For each mutant line, forward

and reverse sequencing reads were aligned to the reference sequence of *met1* using Mega7 (Kumar *et al.* 2016) to allow correct identification of the variants.

The same method was used subsequently to follow the mutations in the progeny of these plants and when crossing them into a clean background.

5.2.12. Barley crossings of *met1* mutants

Seeds from TILLING mutant lines of interest and from *cv.* Barke and Golden Promise were sown into large 8-inch round pots with Cereal Mix (added insecticide) and grown under controlled greenhouse conditions (16 hours day, 8 hours night) for 7 to 8 weeks with regular watering. When the plants were mature, female plants were emasculated by dissecting where the spike is situated and removing all three anthers in all the spikelets using fine tweezers. The plants were then left to mature further for another couple of days. When the female plants were ready, male plants were prepared by dissecting the spike out of the stem and cutting the very top of the spikelets. The plants were then left for 15 minutes whilst the anthers extrude on the male plants. The male spike was then cut off the plant and used to pollinate the prepared female plants. After a week, seeds started to form in the spikelets of the female plant. The plants were left to grow until the F_1 seeds were mature and could be harvested (Booth 2016).

5.2.13. Verifying the success of the crosses with Barke

F_1 seeds, which were obtained from the TILLING mutants crossed with Barke, were sown, and grown in large 8-inches round pots under controlled greenhouse conditions (16 hours day, 8 hours night). DNA was extracted from two-leaf plants as described in 2.2.1.1 and PCR was carried out using primers flanking the region where the mutation was present as described in 5.2.11. The PCR products were then cleaned up using a PCR purification kit (2.2.3) and sent to the genomics facility at the James Hutton Institute for Sanger sequencing. The same DNA was then used for large scale genotyping using the 50K iSelect SNP array (Bayer *et al.* 2017) developed within the institute. Genotyping was performed by GeneSeek® (Illumina® UK). Genotyping results were analysed using Flapjack (Milne *et al.* 2010) to check the F_1 seeds were true hybrids.

5.2.14. Genotyping the F_2 generation

F_2 seeds were recovered from F_1 individuals which were confirmed as true hybrids between Barke and Golden Promise and which carried the mutation of interest in *met1*. The seeds were germinated for two weeks in 4-inch square pots with cereal mix under

controlled greenhouse conditions (16 hours day, 8 hours night) as described in 2.1.2. DNA was then extracted by the genomics facility at the James Hutton Institute.

Presence/absence of the mutation was determined by sequencing. Briefly, the amplicons of interest were amplified using PCR as described in 5.2.11. The PCR products were then cleaned using AMPure XP Beads (Beckman Coulter) as described in 5.2.3. The samples were given to the genomics facility at the James Hutton Institute for Sanger sequencing.

5.2.15. Genotyping of F_3 families

Sequencing results from the F_2 plants allowed the separation of the individuals into three categories: homozygous wild type, homozygous mutant and heterozygous. Twelve families carrying the wild-type allele were chosen at random along with twelve families carrying the homozygous mutant allele, totalling 24 F_3 families. For each family, 12 seeds were germinated as described in 2.1.2 and DNA was isolated by the Genomics Facility service housed at the James Hutton Institute from eight two-leaf seedlings chosen randomly. In total, 192 individual plants were genotyped using the 50K iSelect SNP array (Bayer *et al.* 2017) by GeneSeek® (Illumina® UK). The F_2 parents of each family were also genotyped using the 50K SNP array using the DNA isolated in 5.2.14.

The recombination analysis was carried out with the help of Dr. Mikel Arrieta (The James Hutton Institute) using an R script which was developed in house. Briefly, the genotyping data was first filtered and cleaned to remove any markers which failed (more than 5% missing data across all genotyped plants) or were homozygous for the F_1 parents (*cvs*. Barke and Golden promise). This left 10,133 polymorphic markers which were heterozygous for this population and that could be used for the subsequent analysis. Allele calls were replaced by the parental genotype: “a” for the homozygous mutant parent (*cv.* Golden Promise) alleles, “b” for the homozygous wild-type parent (*cv.* Barke) alleles, and “h” for heterozygous markers. Subsequently, F_1-related recombination events were excluded by removing for each family marker regions which were homozygous in their respective F_2 parent. Quality Control of the data was performed ensuring the pedigree of each family was correct by verifying that homozygous regions identified in F_2 plants corresponded to homozygous regions in their corresponding F_3 plants and checking that the area surrounding the mutation of interest matched the homozygous genotype associated to the parental variety. This identified 12 families for which the F_3 plants genotype did not match that of their respective F_2 parent, and for

which the mutation region was heterozygous. This was later explained by an accidental inversion of trays carrying the plants at the F_2 stage (5.2.14). The data from these individuals had to be eliminated from the analysis, resulting in only six families, with a total of 48 plants, per treatment, to be analysed.

During the recombination distance calculation analysis, the marker order was determined by the physical reference map "Morex V2" as published by Monat *et al.* (2019). Crossover (CO) events were determined as allele changes from either homozygous allele call ("a" or "b") to a heterozygous call, or vice-versa. Changes from one homozygous allele call to the other ("a" to "b" or "b" to "a") were counted as a double CO event. Only changes that were maintained in the following three markers were accepted as true CO events, in order to ensure allelic change numbers were not affected by genotyping errors. Genetic distance between markers was calculated using the Kosambi formula (Kosambi 1944). Statistical analysis between the obtained genetic maps was carried out using a Wilcoxon signed-rank test (Wilcoxon 1945). Markers were also attributed to their corresponding genomic region as described by Beier *et al.* (2017).

5.2.16. Identifying semi-sterile *met1* mutants using exome capture data

Identification of Golden Promise mutants in *met1* was carried out using data obtained by Schreiber et *al*. (2019) from Targeted Exome Capture Sequencing performed on 274 semi-sterile M_3* lines. Briefly, sequencing reads were mapped against the Golden Promise reference assembly (Schreiber *et al.* 2019) using bwa mem (v0.7.17). Samtools was then used to filter (alignment score ≤ 70) and sort the alignments. The variant calling was performed using Freebayes, and the variants filtered using Vcftools and Vcffilter. Indels were then removed, and the SNP set was filtered to only keep those with more than 4 reads as support. Finally, only SNPs present in *Met1* were selected from the original dataset and their PROVEAN score calculated as described in 5.2.10.

5.2.17. Identifying *Met1* and *Ddm1* natural variants in wild barley accessions with exome capture data analysis

Similar to 5.2.16, natural variants in *met1* were identified using data published by Russell *et al.* (2016) on 267 landraces and wild barley accessions. Data alignment and variant calling were performed as described in 5.2.16, selecting for SNPs identified in

Met1 and *Ddm1* only, and their respective PROVEAN score was calculated as detailed in 5.2.10. For this project, variants in *met1* were selected.

5.2.18. sgRNA design for CRISPR-Cas9

In order to function properly, Cas9 requires a guide RNA (gRNA) which will specifically target the gene to be mutated. Synthetic guide RNAs (sgRNA) were designed for both *met1* and *ddm1_4H* in order to target the very start of each gene, and preferentially their second exon, with the hope of completely knocking-out the gene. Two different algorithms were used to look for suitable targets within the genes: sgRNA Scorer 2.0 has been developed by the Church lab from the University of Harvard (https://crispr.med.harvard.edu/sgRNAScorer) (Chari *et al.* 2017) and sgRNA Design for CRISPR is a tool from the Broad Institute's Genetic Perturbation Platform in Cambridge, Massachussets (http://www.broadinstitute.org/rnai/public/analysis-tools/sgrna-design) (Sanson *et al.* 2018). The list of obtained sgRNA was then analysed, selecting only the ones that were detected by both algorithms. Both algorithms are optimised for mammalian genomes, and even though both genes are heavily conserved amongst all species (see 5.3.1), confidence in the efficiency of the sgRNA would be higher if detected by two independent methods.

From the list of sgRNA candidates, the ones with the highest confidence score from both algorithms were selected. The sgRNA also had to have a G on their 5' end and a G or an A on their 3' end. The selected sgRNA were then subjected to a BLAST search in the barley genome (http://www.barlex.barleysequence.org) (Colmsee et al., 2015), to ensure their specificity and that there would be no off-target hits, especially for *ddm1_4H*. As this gene is present in three different copies in the genome (see 5.3.1), the sgRNA had to be specific for the 4H version of the gene only.

5.2.19. Making the double sgRNA insert

The aim of the cloning was to generate a construct which contains both sgRNAs. This way, in one single transformation, the transgenic plant would potentially carry both mutations for *met1* and *ddm1_4H* or be a single mutant for either of the genes.

To achieve this goal, primers were designed, which would allow amplification of an insert containing the sgRNA for *met1*, followed by its termination sequence, then the promoter and second sgRNA for *ddm1_4H* using as a template a pre-existing plasmid containing two sgRNA for two other meiotic genes, FIGL-1 and FANCM, which had previously been

designed and used by colleagues at the James Hutton Institute. Amplification of this insert would effectively replace the sgRNAs targeting FIGL-1 and FANCM with the sgRNAs designed to target Met1 and Ddm1_4H.

The primers contain an *AarI* digestion site CACCTGC (4/8)^ where the final 4 nucleotides are complimentary to the sticky ends of the plasmid, followed by the spacer and the 23 first nucleotides of the sgRNA terminator (forward primer) or the 23 last nucleotides of the sgRNA promotor (reverse primer). These oligonucleotides are described in Appendix 8.1.

PCR amplification of the insert used Q5 HotStart HiFi Polymerase as described in 2.2.2.1 and the amplicons were purified as detailed in 2.2.3. Correct amplification of the insert was verified by electrophoresis gel migration (2.2.6).

5.2.20. Subcloning into pGEM®-T Easy

The insert was first cloned into a pGEM®-T easy vector for multiplication purposes, to make digestion at each end of the insert with *AarI* easier and to avoid over-digestion of each end of the insert.

For the insert to be correctly ligated into the vector, single-A overhangs were added to the PCR product on each extremity. The amplified, purified, fragment was incubated at 72°C for 20 minutes in a reaction containing 0.2 mM of dATPs, 1X GoTaq Buffer, 5 U of GoTaq DNA Polymerase and around 1 mg of insert. The reaction was inactivated at 8°C and the insert can be stored until the ligation process.

Ligation into the plasmid was performed using T4 DNA ligase (New England Biolabs) following the supplier's indications. The ligation reaction was performed at 16°C overnight, then stopped by placing the samples at 4°C for 30 minutes.

The ligated vector was then transformed into 10-beta competent *E. coli* cells with high efficiency (New England Biolabs). The cells had previously been thawed on ice for 10 minutes before 50 µL aliquot was transferred to a chilled 15 mL Falcon tube along with 5 µL of ligation mixture. The reaction was then gently mixed before being incubated for 30 minutes on ice. Plasmids were inserted into cells using heat shock in a water bath heated at 42°C for 30 seconds before placing the cells on ice for 5 minutes. 950 µL of pre-heated SOB (Appendix 8.4) at 37°C was then added and the cells were shaken for one hour at 37°C, 300 rpm, for recovery. 100 µL of the recovery culture was then spread on

a pre-heated IX-Amp LB (Appendix 8.4) culture plate at 37°C, which contains 100 µg/mL Ampicillin, 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and 80 µg/µL 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) for selection. Cells which contain the insert are not able to metabolise X-gal and will result in white colonies whereas the ones where the plasmid closed on itself will produce blue colonies. After air-drying, the plates were incubated at 37°C overnight upside-down.

5.2.21. Colony screening

Colony screening was carried out by directly performing a PCR reaction on white colonies which are likely to contain the insert. The PCR reaction was made using GoTaq G2 Flexi DNA polymerase (Promega) and M13 primers (Appendix 8.1) which flank the area in which the insert has been inserted as no polymerase proof-reading capability was needed here.

White colonies were first resuspended in 20 µL SDW and 2 µL of resuspended colonies were used in each reaction, which also contained 1X Green GoTaq Flexi Buffer, 1 mM MgCl2, 0.2 mM dNTPs, 1 mM M13F and M13R primers and 1.25 U GoTaq G2 Flexi DNA Polymerase with a final volume of 25 µL.

The PCR reactions were performed according to the following protocol: initial denaturation at 94°C for 10 minutes; 30 cycles of denaturation at 94°C for 30 seconds, annealing at 54°C for 30 seconds and extension at 72°C for 30 seconds; final extension at 72°C for 5 minutes.

The PCR products were then run on a 1% agarose electrophoresis gel as described in 2.2.6. The resuspended colonies that showed a positive result after colony screening were then re-inoculated in a 15 mL culture containing LB medium with 100 µg/mL ampicillin and grown overnight at 37°C under gentle agitation (250 rpm).

5.2.22. Plasmid purification

Plasmid DNA extraction was completed using the GeneJET Plasmid Miniprep Kit (Thermo Scientific). The liquid cultures were first centrifuged for 15 minutes at 3,000 rpm, at room temperature to pellet all the cells and the rest of the plasmid isolation was carried out following the manufacturers' guidance. Briefly, the pelleted cells were resuspended in 250 µL of resuspension solution and transferred to a fresh 1.5 mL Eppendorf tube. 250 µL of Lysis solution was added and the tubes were inverted 4-6 times and left to stand for

up to 3 minutes at room temperature. 350 µL of Neutralization Solution were then added to the samples, and the tubes were inverted 4-6 times before being centrifuged at 12,000 x g for 5 minutes. The supernatant was then transferred to a GeneJET Spin column and the samples were centrifuged for 1 minute at 12,000 x g, and the flowthrough was discarded. The columns were then washed twice by adding 500 µL of Wash Solution and spinning the samples for 1 minute at 12,000 x g, then removing the flowthrough. The empty columns were then centrifuged for 1 minute at 12,000 x g to remove any excess Wash Solution. The columns were then transferred to a fresh 1.5 mL Eppendorf and 50 µL of Elution Buffer were added to each sample. The columns were left to stand for 2 minutes at room temperature, then were centrifuged for 2 minutes at 12,000 x g. The eluted plasmids were aliquoted and given to the Genomics facility housed at the James Hutton Institute for Sanger Sequencing using Met1-R and Ddm1-F primers described in Appendix 8.1.

5.2.23. Cloning into p-Bract214m-HvCas9-HSPT

After sequencing results confirmed the pGEM®-Teasy vector contained the full insert, the plasmid was cut with the restriction enzyme *AarI* in the following 20 µM reaction: 1X Buffer *AarI*, 500 ng plasmid DNA, 1X provided Oligonucleotides mix and 2 U *AarI* (Thermo Scientific). The digestion was incubated for 2 hours at 37°C. The digestion products were then run on a 1% agarose gel to check that the size of the insert was correct. The band containing the digested insert was then cut out of the gel under UV light and the insert was purified using a QIAquick Gel Extraction Kit (Qiagen) following the supplier's instructions. Briefly, 3 volumes of Buffer QG were added to 1 volume of the excised gel fragment (1 volume = 100-300 µL depending on the sample), and the samples were incubated for 10 minutes at 50°C to allow for the complete dissolution of the gel. 1 volume of isopropanol was then added to each sample and the tubes were inverted 4-6 times to mix the samples. The samples were then transferred to a QiaQuick column and were centrifuged for 1 minute at 17,000 x g. The flowthrough was discarded, and the columns were washed with 750 µL of Buffer PE, and centrifuged for 1 minute at 17,000 x g. After discarding the flowthrough, the empty columns were centrifuged again for 1 minute at 17,000 x g to remove any excess Buffer PE. The column was then placed into a fresh 1.5 mL Eppendorf, and the DNA was eluted by adding 50 µL of Buffer EB (10 mM Tris, pH 8.5), leaving the columns to stand for 1 minute, and centrifugating the

samples at 17,000 x g for 1 minute. DNA concentration was measured using a NanoDrop 2000™ spectrophotometer (ThermoFisher).

Ligation of the insert into the p-Bract214m-HvCas9-HSPT vector was carried out following the protocol described in 5.2.20. The ligation mixture was then transformed into XL-10 Gold Ultracompetent *E. coli* cells (Agilent) following the manufacturers' protocol. In brief, a 1.5 mL Eppendorf containing a 50 µL aliquot of cells was thawed on ice for each transformation reaction. 2 µL of β-mercaptoethanol were added to each tube and the cells were left to incubate on ice for 10 minutes, gently swirling the tubes every 2 minutes. 2 µL of the ligation mixture were added to the tubes, and the cells were mixed gently by flicking the bottom of the tubes before a 30-minute incubation on ice. The transformation reaction was then done by heat shock, incubating the tubes for 30 seconds at 42°C. The cells were immediately transferred back on ice and left to rest for 2 minutes before being resuspended in 950 µL of pre-warmed SOB (37°C). The cells were then left to recover for an hour at 37°C, shaking at 300 rpm. After recovery, the cells were plated onto LB Medium with 50 µg/mL Kanamycin and left to incubate overnight at 37°C. The final vector map containing both sgRNAs for *met1* and *ddm1_4H* is available in Appendix 8.5.

Colony PCR was performed as described in 5.2.21. The positive colonies were re-suspended in a 15 mL liquid culture supplemented with 50 µg/mL Kanamycin and left to incubate overnight at 37°C under gentle agitation. Plasmid purification was then performed following the protocol detailed in 5.2.22.

Integrity of the plasmid was checked by digestion using the restriction enzyme *BglI* (New England Biolabs), which should result in five fragments of 552 bp, 1025 bp, 1974 bp, 2806 bp and 5901 bp: 1 µg of plasmid DNA was incubated with 1 µL of 10X CutSmart Buffer and 0.5 µL of *BglI* (20 U/µL) in a 20 µL reaction. The digestion was left at 37°C overnight and was then reactivated with 10 U *BglI* the subsequent morning for 2 hours before being inactivated 20 minutes at 65°C. Digestion products were then run on a 1.4% agarose gel.

When integrity of the plasmid was confirmed, it was sent for Sanger sequencing along with Met1-F and Ddm1-R primers described in 5.2.22.

5.2.24. Cloning into AGL-1 and embryo transformation

After sequencing confirmed the integrity of the plasmid, it was transformed into *Agrobacterium tumefaciens*, strain AGL-1-pSoup, by electroporation. Briefly, after the cells had been left to thaw on ice, 1 µL of the plasmid was placed with 20 µL of cells in electroporation cuvettes, ensuring extra care was taken so no air bubbles would be present in the cuvettes. Electroporation was carried out in a Gene Pulser Xcell™ by Bio-Rad (200 Ω, 25 µF, 2.5 kV) and the electroporated cells were immediately transferred into 1 mL Yeast Extract Beef (YEB) medium for recovery at 28°C under gentle agitation (200 rpm) for 2 hours. After recovery, 10 µL of cells were spread onto LB plates supplemented with 100 µg/mL Kanamycin and left to incubate for 48 hours at 28°C. From the resulting colonies, screening was completed as described in 5.2.21 on the corresponding colonies to confirm presence of the plasmid, and 15 mL liquid cultures were made in YEB medium (Appendix 8.4) supplemented with 100 µg/mL Kanamycin.

The transformation process was performed accordingly to Bartlett et al. (2008). The early stages of embryo transformation were carried out as part of this PhD thesis, whilst the later stage manipulations were done by the Genetic Transformation Facility on site at the James Hutton Institute. As a positive control, separate embryos were also transformed with an *Agrobacterium* liquid culture containing a vector known to have been successfully transformed in the past at the James Hutton Institute. Similarly, embryos were transformed with empty *Agrobacterium* cells as a negative control.

5.3. Results and Discussion

5.3.1. *In silico* characterisation of *Met1* and *Ddm1* in barley

In plants, two main genes are involved in the maintenance of genome-wide methylation levels: DNA (cytosine-5) Methyltransferase 1 (*Met1*) and Decrease in DNA-methylation 1 (*Ddm1*). Both genes were characterised *in silico* in barley as described in 5.2.1.

*5.3.1.1. DNA (cytosine-5) Methyltransferase 1 (*Met1*)*

The *Met1* gene in *cv*. Golden Promise is situated on chromosome 2H (58 cM). The full gene sequence is 9,945 bp long, and the CDS is 4,608 bp long. The gene is composed of 12 exons (Figure 5.3.1). Gene expression profiles of *Met1*, as recovered from the Barlex Genome Database (Colmsee *et al.* 2015), showed this gene was most highly expressed during the development of inflorescence stages in barley. Prediction of functional domains revealed the protein was likely carrying a Bromo-Adjacent Homology Domain,

which is commonly associated with chromatin-binding proteins (Chambers *et al.* 2013), especially DNA-methyltransferases, and has been found to play a major role in DNA-methylation in mammals (Callebaut *et al.* 1999). The protein also contains a C-5 cytosine-specific DNA-methyltransferase active site, similar to that identified in mammalian DNA-methyltransferases (Chen *et al.* 1991). Finally, alignment of the protein sequence using the Clustal Omega database (Madeira *et al.* 2019) revealed that the protein was extremely well conserved across plant species, with 16% of the amino acids being untouched amongst all analysed species (Supplementary Data 2).

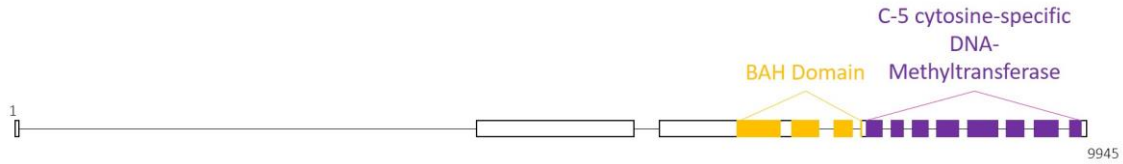


Figure 5.3.1. Gene structure of *Met1* in barley (*cv.* Golden Promise). The gene (9,945 bp) is composed of 12 exons and contains a Bromo-Adjacent Homology (BAH) domain and a C-5 cytosine-specific DNA-Methyltransferase active site.

*5.3.1.2. Decrease in DNA-methylation 1 (*Ddm1*)*

Ddm1 is not a methyltransferase and belongs to the family of SWI/SNF2-like chromatin-remodelling proteins (Jeddeloh *et al.* 1999). Contrary to *Met1*, which is present as a single copy in barley, *Ddm1* is present in three copies in the barley genome, on chromosome 2H (20 cM), 4H (unmapped) and 7H (77 cM). Each gene is 5,692 bp, 6,012 bp and 8,042 bp long, respectively. Their CDS length is relatively similar (2,358 bp, 2,481 bp, 2,364 bp, respectively), however, the *Ddm1* copies situated on chromosomes 2H and 4H contain 14 exons (Figure 5.3.2) whereas the 7H copy only counts 13 exons. Their gene expression profiles show they are predominantly expressed in developing grains (2H), in developing inflorescences (4H and 7H) and in developing tillers (7H). As the 4H copy was predominantly expressed during the inflorescence development stages, where meiosis would occur, only this copy of the gene was used for further analyses. Functional domain prediction highlights the presence of an SNF2-related, N-terminal domain, which is typically present in proteins involved in chromatin unwinding and transcription regulation (Eisen *et al.* 1995) and functions as their ATPase component, allowing for the use of the released ATP energy to disrupt histone-DNA interactions (Dürr *et al.* 2006). This analysis also identified a large helicase, C-terminal domain, which would use the energy liberated from the aforementioned ATP hydrolysis to uncoil DNA molecules, and is traditionally found in members of the SNF2 family (Thomä *et al.* 2005). Similar to *Met1*, *Ddm1* is

extremely well conserved across plant species, with nearly a third of the protein's amino acid being identical between species (Supplementary Data 3).

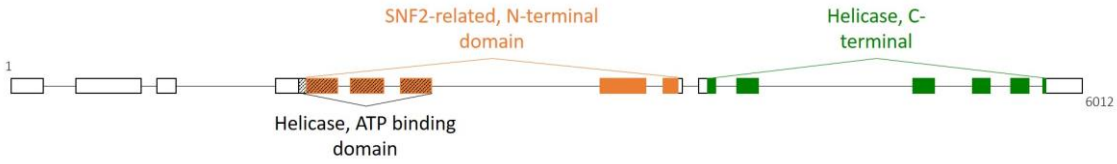


Figure 5.3.2. Gene structure for *Ddm1_4H* in barley (*cv*. Golden Promise). The gene is 6,012 bp long and is composed of 14 exons. It contains an SNF2 ATP binding domain and a helicase active domain.

5.3.2. Detecting randomly induced variations using TILLING

As no known mutant in *Met1* in barley had been characterised at the start of this PhD project, I chose to screen an available TILLING population for mutations in *Met1* in *cv*. Golden Promise. This TILLING population was developed at the James Hutton Institute (Schreiber *et al.* 2019). This population was screened using a single-gene approach, which involved high-throughput multiplexed PCR amplification and sequencing of *Met1* from the two-dimensional DNA-pools mentioned in 5.2.2. This first required PCR optimisation to allow for appropriate amplification of the multiplexed PCR products (5.2.4 and 5.2.5) to allow for the construction of high-quality sequencing libraries for the entirety of the TILLING population. The population was then sequenced for mutations in *Met1* by MiSeq in four batches of 768 plants, and mutation screening was performed by variant calling to identify SNP alleles in the gene.

5.3.2.1. Optimising the library construction

An annealing temperature response experiment was first performed with annealing temperatures ranging from 59°C to 65°C, as described in 2.2.2.1. This experiment was done with each of the primer sets and indicated that the optimal annealing temperature across all primer pairs was 62°C. Determining the optimal annealing conditions for all primer pairs was crucial to the multiplexing aspect of building the sequencing libraries, as all primers will be mixed together in the original PCR amplification step.

Multiplexing the primer pairs was first tested by mixing all primers together in the same mix, as described in 5.2.4. None of the multiplexing reactions resulted in clear PCR product bands at the expected size of 400 bp (Figure 5.3.3). Most primer pairs in the multiplex overlap, which is crucial for extensive screening of mutants in the entirety of the gene. However, this probably causes mix-ups during PCR, where some forward primers might associate with the incorrect reverse primers, or even form secondary structures by complementing one another.

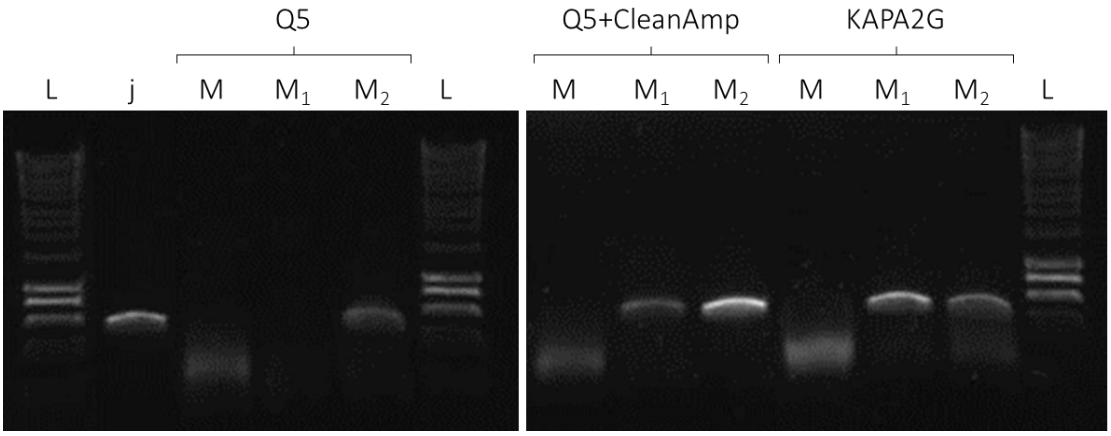


Figure 5.3.3. Test of multiplexing mixes 1 and 2 with non-overlapping pairs of primers (M_1 and M_2), compared to Singleplex amplification (j) and full multiplexing (M). Three conditions were tested: using Q5 polymerase with regular dNTPs (Q5) or with CleanAmp dNTPs (Q5+CleanAmp) and using KAPA2G Multiplex Mastermix (KAPA2G). L: size marker 1 kb+ DNA ladder (NEB)

Two multiplexing mixes, M_1 and M_2, were then created which allowed the overlapping pairs of primers to be separated as described on Figure 5.2.3. When using Q5 polymerase with regular dNTPs, amplification was limited for M_2 and inexistent for M_1 (Figure 5.3.3). However, when coupling Q5 polymerase with CleanAmp dNTPs (Sigma Aldrich, discontinued), a clear band appeared at the expected size of 400 bp for both M_1 and M_2, stronger for the latter. Similarly, a clear band was present for both mixes when using KAPA2G multiplexing mix, however a small smear was also present under the bands, suggesting the creation of by-products and/or primer dimers during amplification which could make sequencing reads harder to analyse. Different extension times were also tested for multiplexing mixes, in order to compensate for the multitude of locations to be amplified within the same sample. PCR products were the cleanest when amplifying for 45 seconds using Q5 polymerase coupled with CleanAmp dNTPs.

Finally, the efficiency of each primer individually within the multiplexing reaction was tested by qPCR as explained in 5.2.7. This allowed for verification that each amplicon was present in the same quantity in the resulting PCR products and that there should be limited differences in the read counts between different regions of the gene. As can be seen in Figure 5.3.4 each primer pair present in the multiplexing mix has amplified at the same rate as the others. Primer pair *n* (Appendix 8.1) was used as a negative control to check for the presence of genomic DNA in the samples. Despite a PCR product clean-up performed as described in 2.2.3, primer pair *n* amplified towards the end of the qPCR reaction, which highlights the importance of a thorough clean-up of the PCR products before proceeding to indexing. For this reason, clean-ups were carried out following the Nextera Illumina protocol for the actual library construction, as described in 5.2.6.

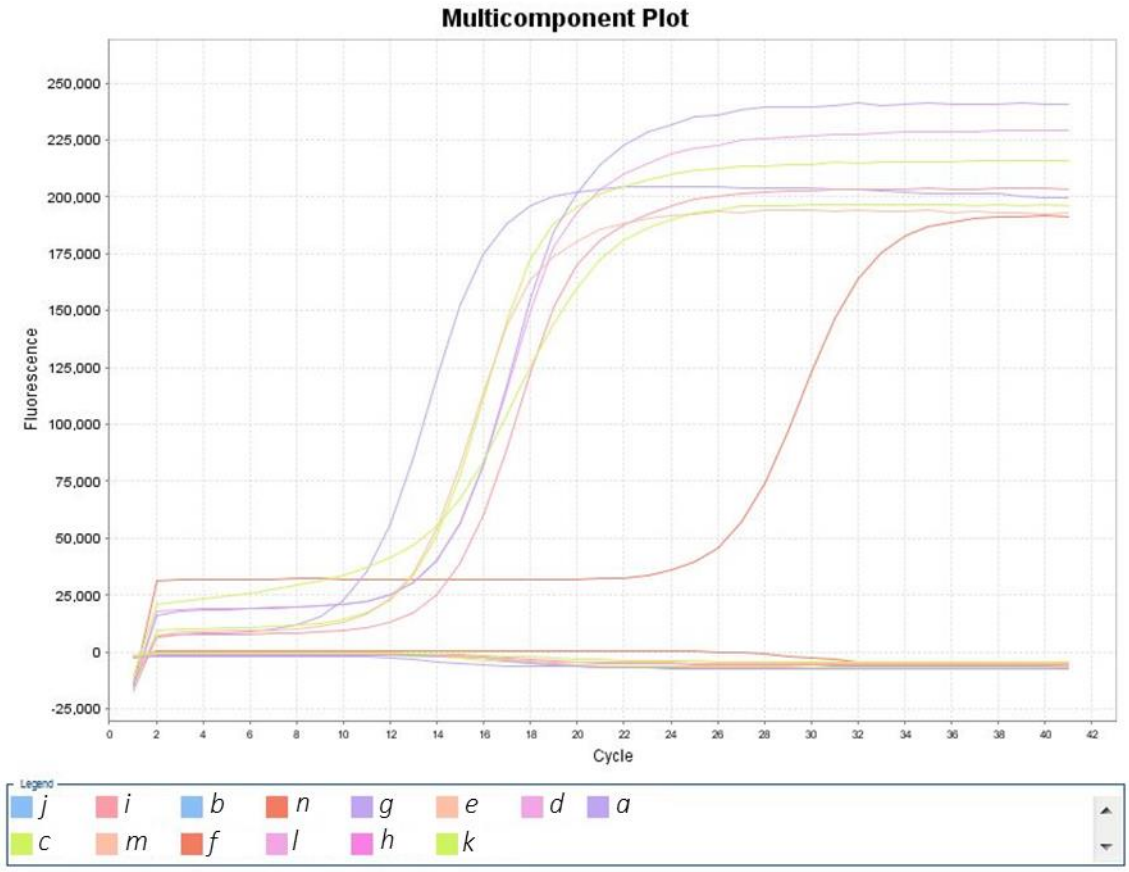


Figure 5.3.4. qPCR amplification curves of each primer pair present in M1 in addition to pair n. The template used was M1 PCR products obtained following protocol 5.2.4.

In order to test whether the indexing process would be efficient for multiplexing, in-house "false-indexing" primers were used to trial a mock indexing reaction, following the protocol described in 5.2.7. These in-house primers consist of the adaptor sequence only of the indexing primers (Appendix 8.1) and were used to avoid using the actual indexing

primers by Illumina which are very expensive and for which quantity was limited to 4 sequencing plates for each primer only.

Tested PCR Products were run on an electrophoresis gel to check the amplicons migrated appropriately. A single band for each sample confirmed that the indexing reaction was performed correctly. These mock samples were cleaned (as per 5.2.8), in order to remove all primers and nucleotides and run on an agarose electrophoresis gel, which confirmed the samples were of an appropriate quality for sequencing.

In order to further check the quality of the optimised sequencing libraries, four mock samples were run on a Bioanalyzer 2100 (Agilent). A single band at 400 bp when running the microfluidic electrophoresis on the Bioanalyzer confirmed the cleanliness of the sample (Figure 5.3.5.a), and the presence of a thin fluorescence peak further confirms the quality of the amplicons and the fact they all amplified at the correct length of 400 bp (Figure 5.3.5.b).

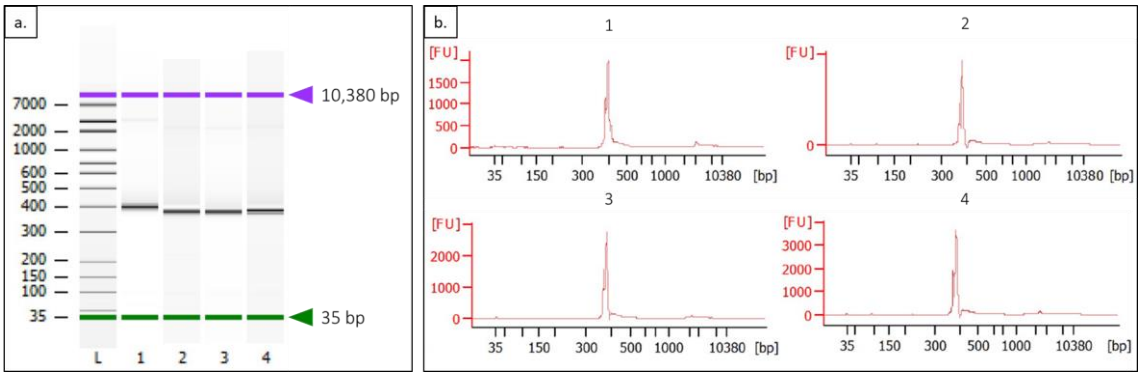


Figure 5.3.5. Bioanalyzer results for final library construction products after false indexing process: a. represents the microfluidic electrophoresis migration of the products; b. is the quantitation of the products going through the electrophoresis process. L: molecular size marker; 1-4: individual samples resulting from multiplexing PCR and false-indexing.

These optimisation steps allowed the design of a comprehensive protocol for building sequencing libraries tailored to the needs of screening for mutations in *met1* by TILLING by sequencing and could both be amended for other target genes and scaled to the whole TILLING population (3,072 individuals, in 4 plates of 768 plants per plate, pooled as described in 5.2.2)

5.3.2.2. Building sequencing libraries

A sequencing library was prepared for each of the DNA plates, containing 768 individuals per plate, as described in 5.2.5, 5.2.6 and 5.2.8. Quality control was carried out at each step by migrating random PCR or cleaned samples from the plate on agarose electrophoresis gels. After the final clean-up, indexed samples were quantified using

Picogreen by the in-house Genomics facility at the James Hutton Institute. This allowed verification that all samples had been processed correctly. Samples with low concentrations were repeated individually and added to the final library in-lieu of the failed sample. The library was then normalised to 4 nM for each sample. All samples were pooled into a single tube following the protocol in 5.2.9. The library was then run on a Bioanalyzer by the James Hutton Institute's Genomic facilities in order to check its quality was satisfactory for MiSeq sequencing (Figure 5.3.6).

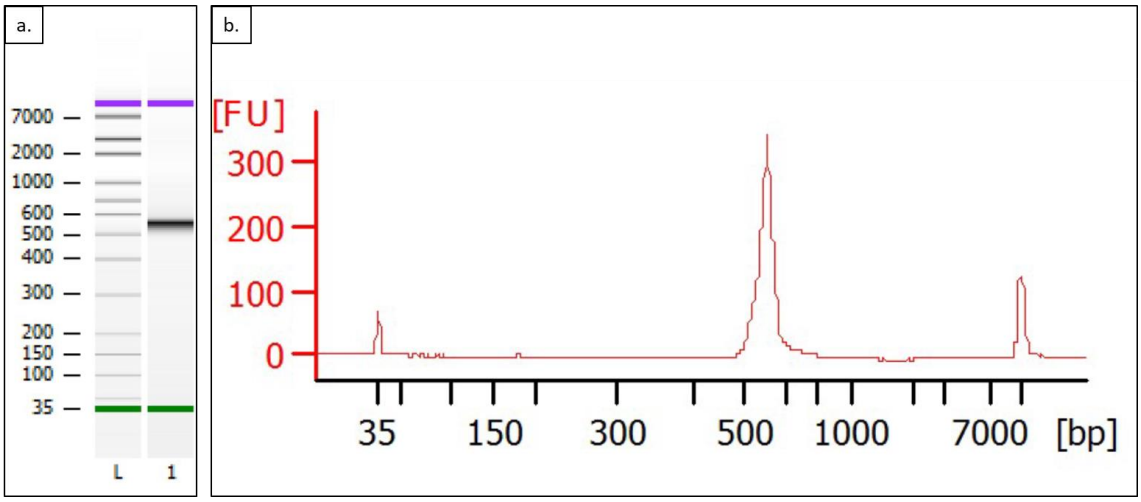


Figure 5.3.6. Bioanalyzer analysis of the final product for the first sequencing library for the first TILLING plate shows a single band between 600 bp and 500 bp as expected. The thin peak at this size indicates the good quality of the library which contains amplicons which are all about the same size.

Once the quality control of each library confirmed sample quality was sufficient for the MiSeq sequencer, the in-house Genomics facilities at the James Hutton Institute performed the sequencing runs and supplied the sequencing reads and data quality analysis.

5.3.2.3. Analysing sequencing data

After sequencing, reads were analysed and aligned on the *met1* gene sequence as described in 5.2.10. The aligned reads were used for variant calling in order to identify mutations caused by the EMS treatment, and the mutations linked back to their plant of origin. In order to facilitate the analysis and avoid the risk of false positives, only point mutations that correspond to the mode of action of EMS (mutating G into A and C into T) were kept. This limited the number of identified mutants, facilitating validation and maintenance of the identified lines. This strategy resulted in the identification of 30 plants carrying 23 different mutations in *met1* within the 3,072 plants screened (Table 5.3.1).

None of these mutations resulted in the formation of a stop codon within the gene. This may be explained by the fact that such a strong mutation could lead to the protein being fully dysfunctional, which could mean the seeds carrying such a mutation would not be able to germinate, let alone generate viable plants, even at a heterozygous state. It is also possible, given the random nature of EMS mutation, that none of the amino acids which could have been turned into a stop codon were affected by the chemical. Despite the absence of stop codons, 18 non-synonymous mutations were identified. This means the change in nucleotide in the gene leads to a codon change and a switch in the resulting amino acid during the translation process. Amongst these, four mutations were predicted to cause a highly deleterious effect on the protein's function by their Provean Score (5.2.10). The Provean score is calculated according to the effect of the mutation on the chemistry of the protein (*e.g.*, changes from an amino acid with a positively charged to a negatively charged chain, from a polar to a non-polar side chain, or changes concerning aromatic amino acids) and the conservation of the concerned amino acid among a wide range of species. The lower the score is, the higher the probability the mutation will cause a strong effect on the protein's function. This list constituted a very promising pool of mutants, which could potentially lead to interesting DNA methylation, and subsequently meiotic and recombinational phenotypes.

Table 5.3.1. List of identified *met1* mutants in the TILLING population. 30 mutant plants were identified which carry 23 different mutations in total: 3 intron mutations, 5 synonymous mutations and 18 non-synonymous mutations. In bold are highlighted the mutations which Provean score's is lower than -5.

Plant No.	Reference	Alternative	Nucleotide position	Amino acid reference	Amino acid alternative	Amino acid position	Amino acid effect	Provean Score
B1C10RE	G	A	2520	Glu	Glu	840	Synonymous	0
B1C10RJ	G	A	Intron	-	-	-	-	-
B1C13RK	G	A	2795	Gly	Asp	932	Nonsynonymous	-1.412
B2C1RG	G	A	2547	Lys	Lys	849	Synonymous	0
B2C4RM	G	A	3967	Ala	Thr	1323	Nonsynonymous	-3.993
B2C8RB	C	T	2164	Pro	Leu	722	Nonsynonymous	-3.007
B2C13RM	G	A	3977	Ser	Asn	1326	Nonsynonymous	-2.995
B2C16RN	C	T	Intron	-	-	-	-	-
B3C3RO	G	A	3358	Ala	Thr	1120	Nonsynonymous	-3.887
B3C5RP	G	A	2884	Val	Ile	962	Nonsynonymous	-0.718
B3C9RA	C	T	2982	Tyr	Tyr	994	Synonymous	0
B3C12RF	G	A	3977	Ser	Asn	1326	Nonsynonymous	-2.995
B3C14RH	C	T	2164	Pro	Leu	722	Nonsynonymous	-3.007
B3C16RG	**C**	**T**	**3994**	**Pro**	**Ser**	**1332**	**Nonsynonymous**	**-7.620**
B5C1RO	C	T	3518	Ala	Val	1173	Nonsynonymous	-3.544
B5C8RD	**C**	**T**	**2618**	**Pro**	**Leu**	**873**	**Nonsynonymous**	**-8.622**
B6C9RF	**C**	**T**	**2618**	**Pro**	**Leu**	**873**	**Nonsynonymous**	**-8.622**
B7C3RB	**C**	**T**	**2618**	**Pro**	**Leu**	**873**	**Nonsynonymous**	**-8.622**
B7C9RP	C	T	4130	Ala	Val	1377	Nonsynonymous	0.976
B7C15RL	C	T	4295	Cys	Thr	1432	Nonsynonymous	-3.757
B8C1RP	C	T	3131	Ala	Val	1044	Nonsynonymous	-2.842
B8C12RJ	**C**	**T**	**4285**	**Pro**	**Ser**	**1429**	**Nonsynonymous**	**-7.781**
B9C7RB	**C**	**T**	**2618**	**Pro**	**Leu**	**873**	**Nonsynonymous**	**-8.622**
B10C11RD	C	T	4295	Cys	Thr	1432	Nonsynonymous	-3.757
B11C1RA	**C**	**T**	**2618**	**Pro**	**Leu**	**873**	**Nonsynonymous**	**-8.622**
B11C1RL	C	T	Intron	-	-	-	-	-
B11C10RE	C	T	2870	Ala	Val	957	Nonsynonymous	-1.430
B12C1RB	C	T	4401	Cys	Cys	1467	Synonymous	0
B12C3RL	G	A	198	Lys	Lys	66	Synonymous	0
B12C9RP	**C**	**T**	**3904**	**Pro**	**Ala**	**1302**	**Nonsynonymous**	**-7.811**

In Figure 5.3.7, most of the identified mutations concern the two domains of interest, especially the active DNA-methylation domain. Moreover, the mutations, which are predicted to have the strongest effect on the protein's function by their Provean score, are all situated in those domains, with 3 associated with the C-5 cytosine-specific DNA-Methyltransferase domain. This highlights the high potential for these mutant lines to produce a faulty version of *met1* and be characterised by a demethylated genome.

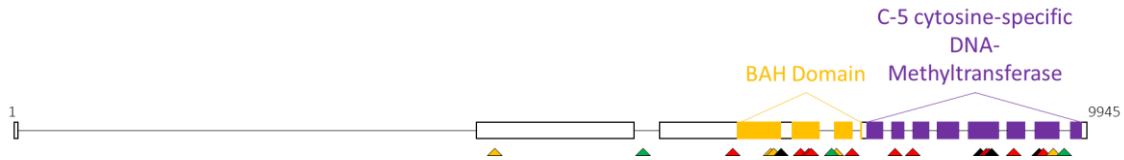


Figure 5.3.7. Location of the identified variants on *met1* show most of the identified mutations concern conserved domains of interest. Mutations are labelled with coloured arrows as follows: green for intron variants, orange for synonymous variants, red for non-synonymous variants with higher Provean scores (> -5) and black for non-synonymous variants with lower Provean scores (< -5).

5.3.2.4. Validating the mutants

To make sure the mutations identified in the M_3* plants were not false positive, M_4* seeds from the selected mutants (Table 5.3.2) were germinated and grown, with a maximum of eight plants per line, when possible. For some of the identified mutant plants, no seed was available, the M_4* seeds would not germinate, or the seed set was too low to be able to sow eight seeds for the line. Validation of the presence of the mutations was performed on these plants as described in 5.2.11.

Mutants from the first plate for which plant material was available were all validated (Table 5.3.2) with the exception of B3C5RP, for which all the sequenced plants were homozygous wild type. For this line, only three plants were obtained, therefore either the mutation was highly deleterious, or the plants were affected by other mutations caused by the mutagenesis which prevented them from germinating. It is also possible that this mutation was falsely identified by the variant calling process. However, considering all the other mutations were validated when seeds were available, this seems improbable.

Table 5.3.2. Details of the validation process for mutants from the first sequencing plate, from block 1 to block 3. Lines for which no plant was sequenced represent lines where no M4 seed set was available, or none of the seeds germinated. B3C5RP could not be validated as none of the sequenced plants carried the mutation. In bold are the non-synonymous lines which were taken further for crossings.

Line	Change	Protein Position	Amino acid reference	Amino acid alternative	Number of plants sequenced	WT/Het/ Mut/NA	Validated?
B1C10RE	Synonymous	840	E	E	5	0/0/5/0	Yes
B1C10RJ	Intron/UTR	-	-	-	3	0/1/2/0	Yes
B1C13RK	**Non-synonymous**	**932**	**G**	**D**	**5**	**2/1/1/1**	**Yes**
B2C13RM	**Non-synonymous**	**1326**	**S**	**N**	**8**	**0/0/8/0**	**Yes**
B2C16RN	Intron/UTR	-	-	-	7	4/0/3/0	Yes
B2C1RG	Synonymous	849	K	K	7	0/0/7/0	Yes
B2C4RM	**Non-synonymous**	**1323**	**A**	**T**	**6**	**0/1/5/0**	**Yes**
B2C8RB	Non-synonymous	722	P	L	0	-	-
B3C12RF	**Non-synonymous**	**1326**	**S**	**N**	**2**	**0/0/2/0**	**Yes**
B3C14RH	Non-synonymous	722	P	L	0	-	-
B3C16RG	**Non-synonymous**	**1332**	**P**	**S**	**7**	**3/0/4/0**	**Yes**
B3C3RO	**Non-synonymous**	**1120**	**A**	**T**	**5**	**2/2/1/0**	**Yes**
B3C5RP	Non-synonymous	962	V	I	3	3/0/0/0	No
B3C9RA	Synonymous	994	Y	Y	0	-	-

Focusing only on Plate 1 (which was screened much earlier than the others), this validation allowed the identification of five viable lines with independent non-synonymous mutations. These mutations are of interest as they are the most likely to influence the protein's function and therefore change overall DNA-methylation patterns and recombination rates in plants (Figure 5.3.8).

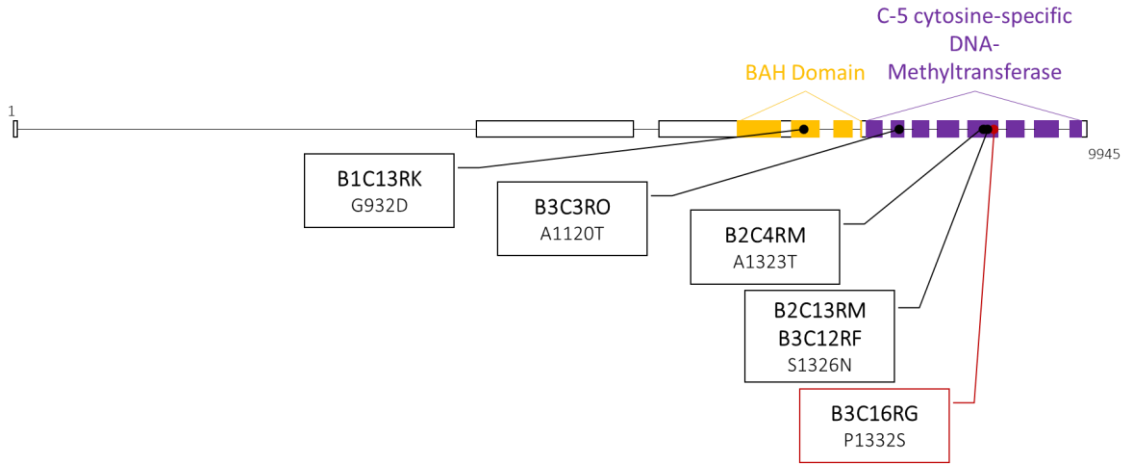


Figure 5.3.8. Location of the mutations for the 6 lines carrying non-synonymous mutations validated from the first sequencing plate. Framed in red is the mutation which is predicted to be highly deleterious by its Provean score (Table 5.3.2). For ease of understanding, lines were renamed after the mutation they carry as explained in each box. For example, B1C13RK is characterized by a change from a Glycine (G) to an Aspartic acid (D) on the 932nd amino acid and was thus renamed G932D.

5.3.2.5. Crossing selected mutants to Barke

M_5* seeds from homozygous M_4* plants were germinated each week for four weeks, in order to produce plants for crossing to *cv*. Barke. If no homozygous seeds were available, heterozygous plants were grown and sequenced as described in 5.2.12. to identify plants carrying the mutations in a homozygous or heterozygous state. This was the case for lines G932D and A1120T, for which homozygous M_4* plants did not produce any seed. Barke plants were also sown every week for 4 weeks, in order to obtain a constant supply of plants to perform crosses. Genotypes containing mutations were directly crossed with Barke since it was nearly impossible to perform back-crosses into Golden Promise and to speed up the process of getting recombination data by F_3 family genotyping.

It proved difficult to emasculate the Golden Promise mutants as the spikelets with their anthers removed would dry very quickly. Therefore, Golden Promise plants were used as pollen donors and Barke plants as female plants (5.2.12). This was not ideal in the case of line A1120T for which no homozygous M_5* plants had germinated, as it reduced the chances of transferring the mutant allele to the F_1 generation.

Of the crosses attempted, three set seed, G932DxBarke, A1120TxBarke and P1332SxBarke. Putative F_1 seeds were recovered, conditioned, and germinated, and after 2 weeks, leaf material was collected, and DNA extracted as described in 2.2.1.1. The presence or absence of the mutations was then checked as described in 5.2.11. As can be seen in Table 5.3.3, none of the F_1 seeds produced by the cross G932DxBarke carried

the mutation. This meant that either the germinated seeds were the result of self-fertilisation, or that only pollen carrying the wild type allele could fertilise the female Barke plants. Only one viable seed was obtained from A1120TxBarke. Luckily, the viable progeny plant carried the original mutation from Golden Promise, despite the M_5* plant being originally heterozygous. Two of the P1332SxBarke progeny carried the mutation in a heterozygous state.

Table 5.3.3. Summary of the genotyping data for the presence/absence of the mutations of interest in the F_1 plants obtained from crosses between the Golden Promise TILLING mutants and Barke.

Plant Number	TILLING Parent Genotype	Crossing Parent	F1 genotype	Successfully crossed?
G932D.1	G/A	G	G	No
G932D.2	G/A	G	G	No
G932D.3	G/A	G	G	No
A1120T.1	G/A	G	G/A	Yes
P1332S.1	A	G	A	Contamination?
P1332S.2	A	G	G/A	Yes
P1332S.3	A	G	G/A	Yes

The plants from lines A1120T and P1332S were then genotyped using the 50K SNP chip (5.2.13) to confirm that they were true F_1 plants. For the three plants, which had been previously identified as true F_1 hybrids, 87% of the SNP markers were heterozygous, confirming the plants were the product of a cross between the Golden Promise TILLING mutants and Barke (Figure 5.3.9). The remaining 13% of markers, which appeared to be homozygous either for the Golden Promise Genotype (7%) or the Barke genotype (6%) can be explained either by genotyping errors or by the SNP markers' allele wrongfully attributed to each genotype in the first place.

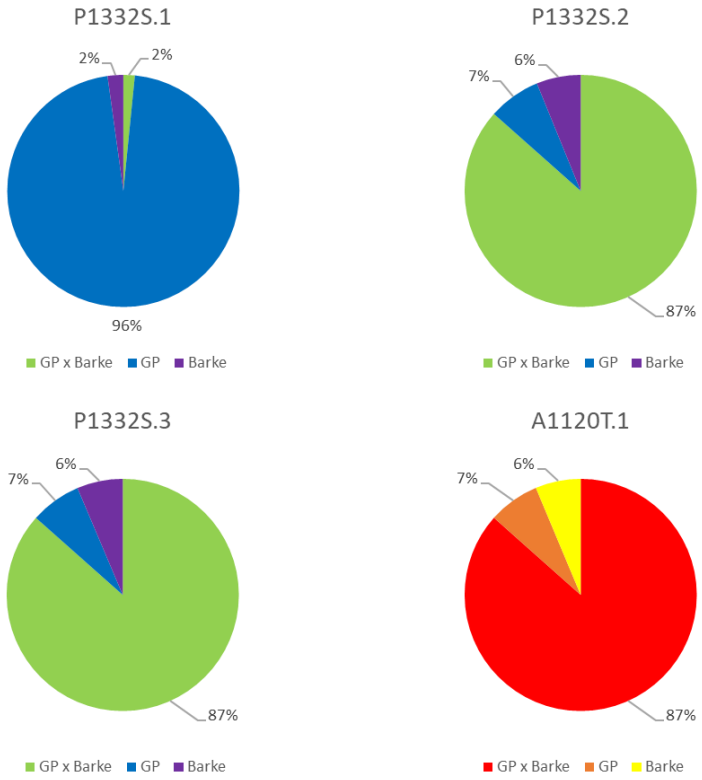


Figure 5.3.9. Distribution of SNP markers in potential F_1 plants shows plants P1332S.2, P1332S.3 and A1120T.1 are characterised by 87% of their markers being heterozygous for Golden Promise and Barke alleles. On the other hand, P1332S.1 sees most of its marker to be homozygous for the Golden Promise alleles, hinting that the plant was a contaminant. Number of SNPs tested: 11,982

5.3.2.6. Analysing F_2 plants

Given the results in 5.3.15, the F_1 lines produced from plants A1120T.1 and P1332S.3 were selected for future use. Their F_2 seeds were harvested, conditioned and germinated (5.5.14) in 8cm pots in 12x8 plants matrices. Seeds which did not germinate were replaced by seedlings from other trays at random in order to obtain a total of 96 plants for each line. After 2 weeks, plant material was harvested and used to assess the presence/absence of the mutations of interest in the F_2 generation as detailed in 5.2.14.

As shown in Figure 5.3.10, the ratio of wild type homozygous, heterozygous, and mutant homozygous alleles was not statistically different than the expected 1:2:1 ratio from a F_2 population ($\chi^2 = 1.61$ for A1120T, $\chi^2 = 4.39$ for P1332S, n=95). Thus, segregation of the mutant and wild type alleles was as expected from a true F_1 hybrid.

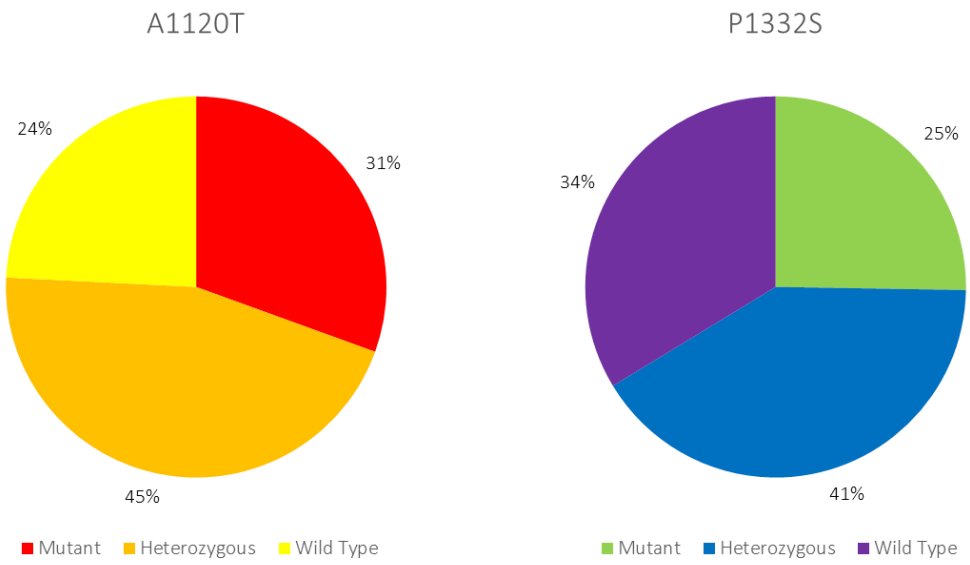


Figure 5.3.10. Distribution of the wild type and mutant alleles in lines A1120T and P1332S are not statistically different from the 2:1:1 ratio expected in an F2 population (n=95).

5.3.2.7. Phenotypes

For each mutation, the number of tillers and plant height were assessed in homozygous mutant, homozygous wild type, and heterozygous plants. As can be seen on Figure 5.3.11, no correlation was found between the presence of the A1120T and P1332S mutations in plants and the number of tillers (A. and B.) or plant height (C. and D,). This suggests the mutations do not affect the function of the *Met1* gene in a sufficient manner to cause severe phenotypes on the plants. Given the severe nature of *met1* mutations observed in rice (Hu *et al.* 2014) and *Arabidopsis* (Mirouze *et al.* 2009), this is of major importance to guarantee fertility and easier maintenance of the mutant lines for further use and study.

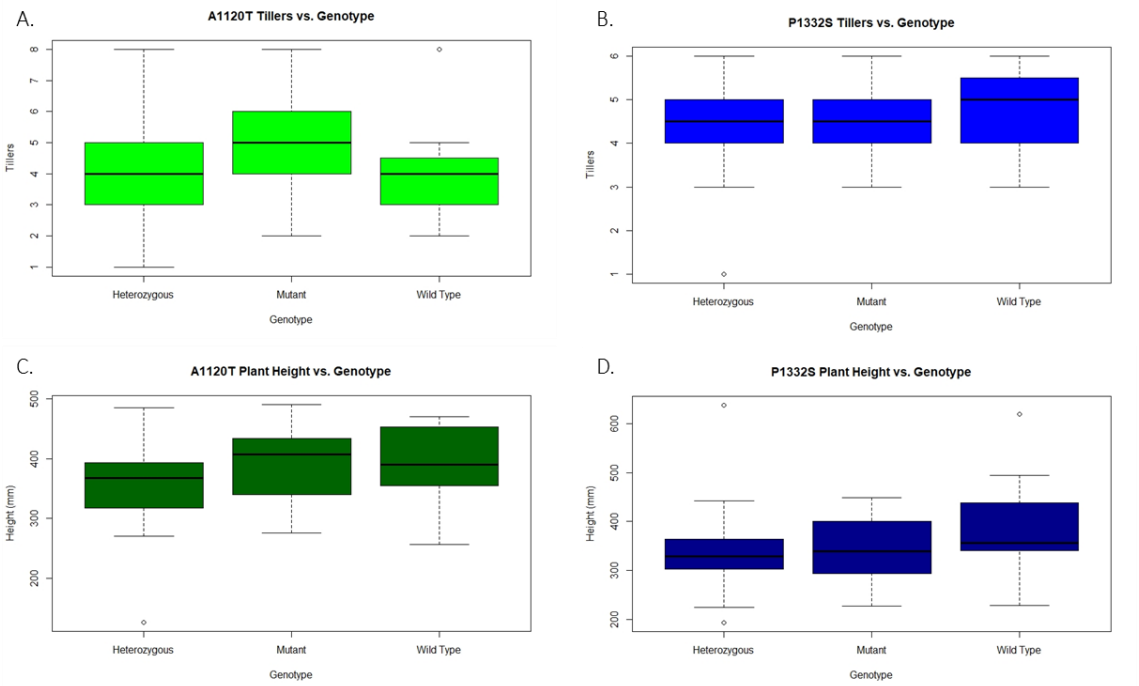


Figure 5.3.11. Phenotypes measured in F_2 plants genotyped as heterozygous, mutant, or wild type for alleles A1120T and P1332S. The presence of the allele does not influence the number of tillers or the height of the plants in A1120T (A. and C.) or P1332S (B. and D.).

5.3.2.8. Recombination analysis in F_3 families

Recombination patterns were analysed in F_3 families resulting from homozygous wild-type or *met1* mutant F_2 parents, using the 50K iSelect SNP array (Bayer *et al.* 2017) as described in 5.2.15. The analysis was done on 48 individuals divided in 6 families for each genotype (WT or *met1*) and used just over 10,000 SNP markers distributed across all 7 chromosomes.

As can be seen on Figure 5.3.12.A, the distribution of crossover (CO) events varied across the different chromosomes, between *met1* and WT plants. On chromosome 1H, 2H and 6H, there seems to be an increase in recombination rates in *met1* mutant plants, especially on the short arm of chromosome 6H. On the contrary, chromosomes 5H and 7H show a tendency to a decrease in recombination events in *met1* mutants compared to WT plants. Chromosome 4H was characterised by a full overlap of the recombination patterns in *met1* and WT plants. Chromosome 3H exhibits a complete lack of heterozygosity in the pericentromeric region between the parental alleles (*cvs*. Golden Promise and Barke), making it virtually impossible to determine whether the *met1* mutation has an effect on meiotic recombination on this particular chromosome. The variability in CO numbers across the whole genome make the difference in recombination frequency non-significant between *met1* plants and WT individuals (Figure 5.3.12.B).

Such differences are possibly a reflection of the low number of individuals in each population (WT and *met1*), which leads to the numbers of individuals being heterozygous for a particular marker to be low and variable. This causes CO events identified in regions with lower numbers of heterozygous individuals to have a higher impact on the genetic maps compared to regions where the number of heterozygous individuals is more consequent, leading to artefacts in recombination frequency across the genome. Despite being non-significant, there seems to be a slight increase in the recombination frequency in *met1* mutants compared to WT plants (Figure 5.3.12.B). This difference is emphasised after dividing the genome into the 3 partitional zones as described by Beier *et al.* (2017), leading to a non-significant increase in CO frequency in zones 1 and 2, and a small decrease in zone 3 (Figure 5.3.12.C). Such observations are aligned with what was previously observed in *met1* and *ddm1* mutants in *Arabidopsis* (Melamed-Bessudo & Levy 2012, Yelina *et al.* 2012), where meiotic recombination rates were increased in sub-telomeric intervals but decreased in peri-centromeric regions. Similarly, in wheat, VIGS silencing of DNA-methylation genes *met1* and *ddm1* was accompanied with elevated recombination frequency in sub-telomeric regions of Chromosome 1A (Raz *et al.* 2021). This tendency is also observed when analysing the distribution of recombination frequency across the relative position of the markers on the genome (Figure 5.3.12.D). There, CO frequency appears to be marginally increased in *met1* mutants on the short arms of the chromosomes, as well as on the long arms of the chromosomes (between 80% and 90%). Wild-type plants present a peak of recombination at 90-95% of the genome, which is likely an artefact caused by the low number of individuals in the population at this position.

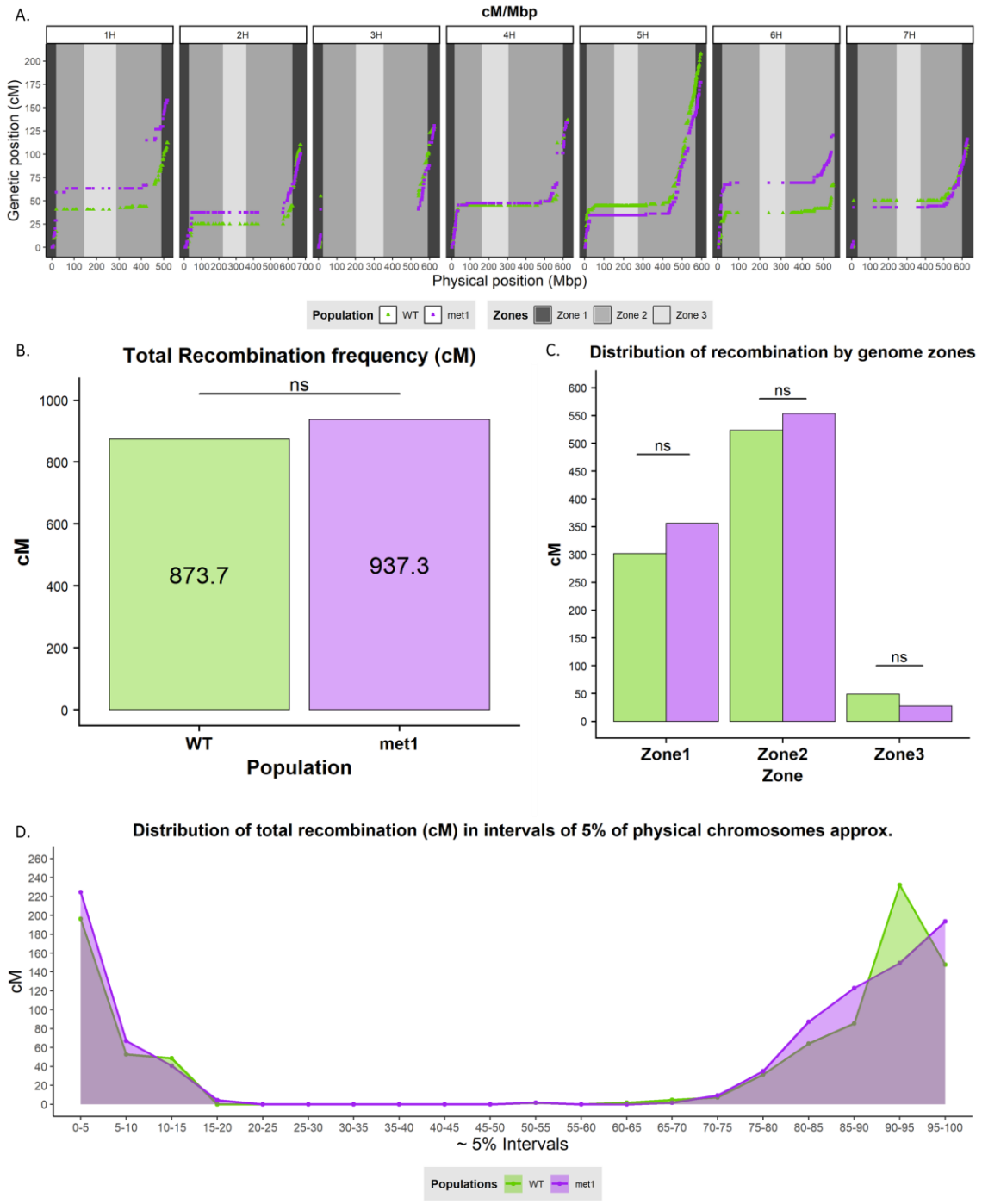


Figure 5.3.12. Analysis of the meiotic recombination landscape in *met1* homozygous mutants (P1332S) compared to wild-type (WT) plants in F_3 families. A: Genetic vs. physical map of the SNP markers for *met1* and WT plants on the 7 barley chromosomes. The effect of the mutation on recombination is variable between chromosomes. Chromosome 3H is characterised by a large region of homozygosity between *cvs*. Barke and Golden Promise. B: Total recombination frequency (in cM) in *met1* mutants and WT plants. There is a slight non-significant increase in the total recombination frequency in met1 mutants. C: Distribution of recombination frequency (in cM) across genome zones as determined by Beier *et al*. (2017) shows a slight non-significant increase in CO frequency in sub-telomeric and distal regions, and a slight non-significant decrease in peri-centromeric regions of the chromosome, for *met1* mutants. D: Distribution on total recombination frequency (in cM) along the relative averaged chromosome position (5% intervals) show a non-significant increase in *met1* mutants in the short arms of chromosomes. There is an inconsistent non-significant difference on the long arm of the chromosomes characterised by a peak at 90-95% in recombination in WT plants. Green: WT plants; Purple: *met1* mutants.

Overall, this analysis shows no statistically significant effect of the *met1* mutation in P1332S on meiotic recombination frequency. However, with the low number of individuals in each condition, only large effects would have been detected. Considering the nature of the mutation, a SNP, and the hypothesis that such a mutation would not cause as strong a phenotype as a complete knockout of the gene, only a mild effect of the mutation on meiotic recombination would be expected (and desired!). In order to verify whether this mild effect was real, the genotypic analysis should be extended to a larger number of individuals, allowing any small effect on the meiotic recombination landscape to be confirmed.

5.3.3. Semi-sterile Golden Promise TILLING mutants with variations in *Met1*

In parallel to screening for *met1* mutants within the TILLING population by a single-gene sequencing (5.3.2), mutations in *met1* were also identified by a reverse-genetic approach in semi-sterile TILLING mutants, as described in 5.2.16. Two lines were identified which carried non-synonymous mutations in *ddm1*, and four lines containing *met1* mutations (Table 5.3.4). None of the mutations identified in *ddm1* were predicted to have a highly deleterious effect by their Provean scores. Lines IP-153 and IP-174 both carry mutations in *met1* which are predicted to have a strong effect on the protein's function.

Table 5.3.4. List of semi-sterile mutants identified by exome capture as carrying non-synonymous mutations for *met1* and *ddm1*.

Line	Gene	Reference	Alternative	Nucleotide position	Amino acid reference	Amino acid alternative	Amino acid position	Change	Provean Score
IP-239	HvDDM1	G	T	459	Lys	Asn	153	nonsynonymous	1.533
IP-176	HvDDM1	C	T	1406	Thr	Ile	469	nonsynonymous	-0.287
IP-153	**HvMET1**	**C**	**T**	**316**	**Pro**	**Ser**	**106**	**nonsynonymous**	**-6.802**
IP-174	**HvMET1**	**C**	**T**	**670**	**Pro**	**Ser**	**224**	**nonsynonymous**	**-6.57**
IP-086	HvMET1	G	A	3763	Ala	Thr	1255	nonsynonymous	3.588
IP-200	HvMET1	G	A	3977	Ser	Asn	1326	nonsynonymous	-2.951

Two of the mutations identified in *met1* are present in the C-5 cytosine-specific DNA-Methyltransferase domain (Figure 5.3.13). Surprisingly, the two mutations, which are predicted to have the strongest effect on the protein's function, are present in the first

half of the second exon, which does not contain either of the strongly conserved functional domains. This could mean these amino acids are involved in protein folding or in recognising where on the DNA the methyltransferase binds to methylate cytosines.

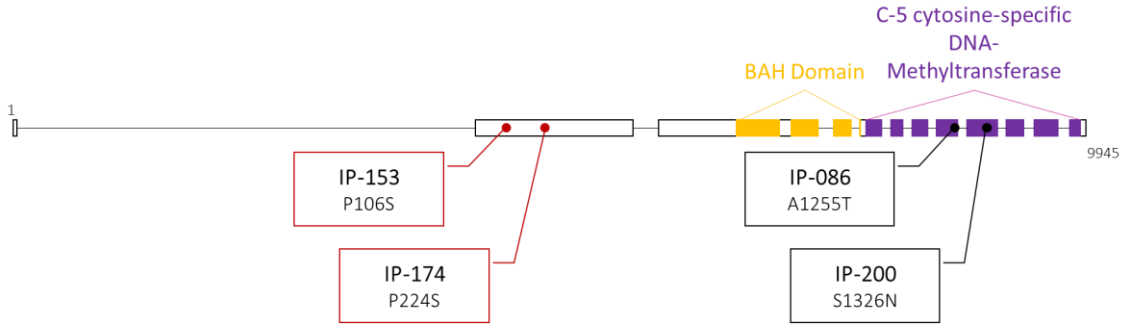


Figure 5.3.13. Position of the *met1* mutations identified in semi-sterile TILLING mutants. Mutations boxed in red are characterised by a Provean score inferior to -5. Similar to Figure 5.3.17, lines (starting with IP-) were renamed depending on the effect of the identified mutation. For example, IP-153 was renamed P106S as the mutation caused a change from a Proline (P) to a Serine (S) on the 106th amino acid.

Only one of the mutations identified in *ddm1* was situated in a domain of interest, with the other located on the second exon of the protein (Figure 5.3.14).

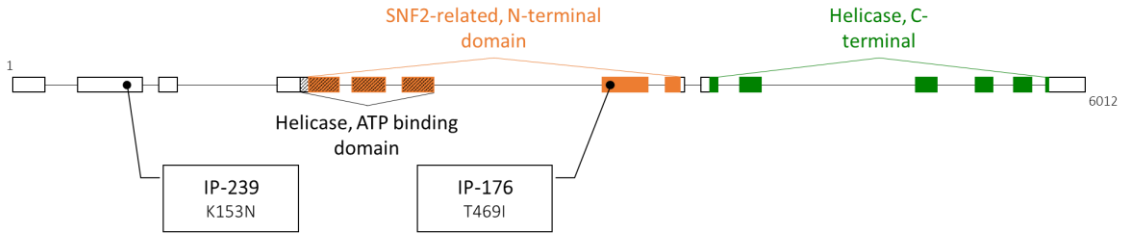


Figure 5.3.14. Position of the *ddm1* mutations identified in semi-sterile TILLING mutants. Again, the lines were renamed according to the effect of the mutation as described in Figure 5.3.22.

In a similar manner to the mutants identified in 5.3.1, the presence of these mutations was verified as described in 5.2.11. Mutations identified in *ddm1* were validated in both lines IP-239 and IP-176 (Table 5.3.5). However, only lines IP-174 and IP-200 (mutations in *met1*) had sufficient seed for validation as mutant lines. The absence of seeds for IP-153 could be possibly explained by the predicted highly deleterious impact of the mutation on protein function (Provean score -6.802, Table 5.3.4). The predicted effect of the mutation carried by line IP-086 was not as deleterious, but its location at the core of the active DNA-methyltransferase domain could have affected plant viability.

Table 5.3.5. Details of the validation process for semi-sterile mutants. Lines where no plants were sequenced did not produce a viable seed set and could not be checked.

Line	Gene	Change	Protein Position	Amino acid reference	Amino acid alternative	Number of plants sequenced	WT/Het/ Mut/NA	Validated?
IP-239	HvDDM1	Non-synonymous	153	K	N	1	0/0/1/0	Yes
IP-176	HvDDM1	Non-synonymous	469	T	I	3	0/0/3/0	Yes
IP-153	HvMET1	Non-synonymous	106	P	S	0	-	-
IP-174	HvMET1	Non-synonymous	224	P	S	4	0/0/4/0	Yes
IP-086	HvMET1	Non-synonymous	1255	A	T	0	-	-
IP-200	HvMET1	Non-synonymous	1326	S	N	3	0/1/2/0	Yes

These may be valuable methylation mutants for future use, but due to time restrictions they were not taken further, and my attention focused on the mutations identified by sequencing in 5.3.1.

5.3.4. Natural variation in methylation genes

Landraces and wild barley cultivars can be an incredible resource for naturally occurring variations in genes of interest. As they are present in fully fertile and viable background genotypes, any detected sequence variants provide the advantage of being isolated from other serious mutations. If the mutation naturally occurs in the wild, it likely only has a mild effect and so maintaining homozygous lines is straightforward. They could of course maintain an altered pattern of DNA-methylation and consequently, recombination landscape.

As described in 5.2.17, pre-existing exome capture data was used to identify variants in *met1* and *ddm1*. A total of 56 sequence variants were identified for *ddm1* in 169 different lines, and 37 mutations were identified for *met1* in 219 lines (Supplementary Table 4). Table 5.3.6 groups the non-synonymous mutations for which the Provean score was low enough for the mutant allele to have a predicted deleterious effect on protein function. This reduced the list to four potentially deleterious mutations in *ddm1* and four in *met1*.

Table 5.3.6. Identification of naturally occurring mutations in *met1* and *ddm1* using exome capture. Only non-synonymous mutations with a Provean Score lower than -2.5 were selected for validation. The complete list is detailed in Supplementary Table 4.

Line	Gene	Reference	Alternative	Nucleotide position	Amino acid reference	Amino acid alternative	Amino acid position	Change	Provean Score
FT584	HvDDM1A	G	T	674	A	E	225	nonsynonymous	-2.87
FT746	**HvDDM1A**	**G**	**T**	**758**	**A**	**D**	**253**	**nonsynonymous**	**-3.61**
FT660	HvDDM1A	T	C	1066	K	E	356	nonsynonymous	-2.89
FT572	**HvDDM1A**	**T**	**G**	**2297**	**D**	**A**	**766**	**nonsynonymous**	**-3.26**
BCC21 BCC51	**HvMET1A**	**T**	**A**	**452**	**E**	**V**	**151**	**nonsynonymous**	**-4.61**
BCC126	HvMET1A	C	A	622	V	L	208	nonsynonymous	-1.58
FT175	HvMET1A	C	T	2915	R	Q	972	nonsynonymous	-2.647
FT31	HvMET1A	T	G	4109	K	T	1370	nonsynonymous	-2.901

Two of the identified mutations were present in the BAH domain and methyltransferase domain in *met1*, whilst the two others were present on the first half of the second exon (Figure 5.3.15). The variant with the predicted highest impact on protein function was in the second exon, similar to what was observed in 5.3.2. This reinforces the possibility for this region to be involved either in binding DNA for methylation or in correct folding of the protein.

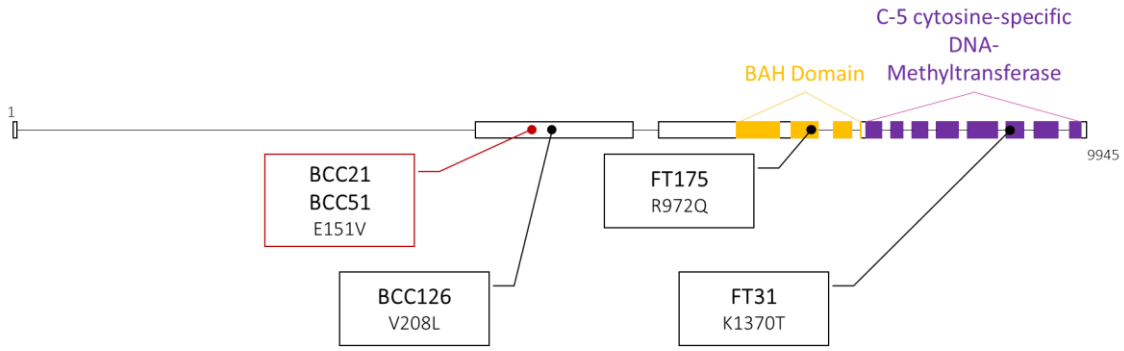


Figure 5.3.15. Overview of the location of the identified natural variants in *met1*. As previously (Figure 5.3.6), the mutation boxed in red represents the one with the lowest Provean score (under -4) and the lines were renamed for clarity depending on the mutation that characterises them. Thusly, BCC21 and BCC51 were renamed E151V as they carry a mutation which transforms the 151st amino acid from a Glutamic acid (E) to a Valine (V).

In the case of mutations in *ddm1*, one of the lines (FT660) carries a variant in the SNF2 active domain, with other mutations situated in the 4th or 14th exon (Figure 5.3.16). The two variants with the lowest PROVEAN score were not in domains of interest. FT746 is characterised by containing a variant situated just upstream from the Helicase ATP binding domain and might be involved in the activity of this domain or in appropriate folding of the protein. The variant carried by FT572 is downstream from the C-terminal Helicase domain and could be involved in the same way either in domain activity or protein folding.

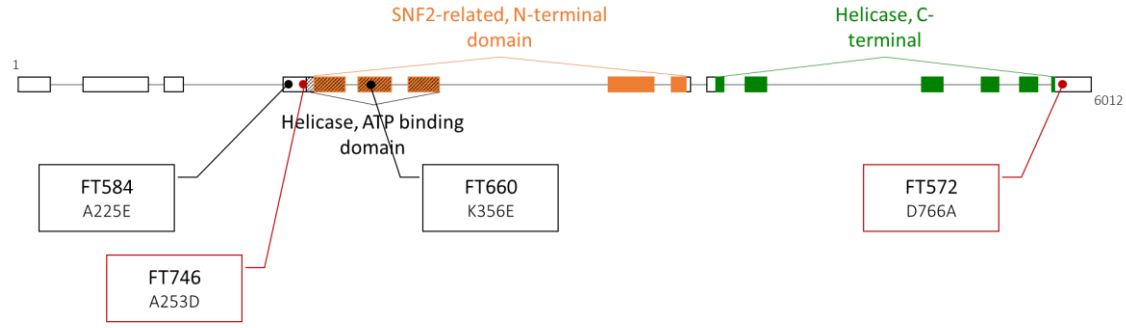


Figure 5.3.16. Mapping of the naturally occurring mutations identified in ddm1. Again, as for Figure 5.3.13, high impact mutations are boxed in red and lines have been renamed to reflect the amino acid change in the protein.

Once again, characterising these variants would require crossing and back-crossing which are both resource and time-demanding. Given the necessary timelines, I decided to prioritise the characterisation of *met1* mutants identified in the TILLING population (5.3.1).

5.3.5. Developing CRISPR mutant lines for *met1* and *ddm1*

Chemically induced random mutations (5.3.2 and 5.3.3) and the identification of natural variants (5.3.4) in *Met1* presented the advantage that many mutations could be generated and isolated in a time and cost-efficient manner. However, the location of the mutations could not be directed to a specific region on the gene, and no stop codons were obtained, meaning the study of mutations with a potential stronger effect on the functionality of the protein could not be done. For this reason, I decided to develop a CRISPR-Cas9 construct with sgRNAs targeting both *met1* and *ddm1* in *cv.* Golden Promise. Transforming barley immature embryos with *Agrobacterium tumefaciens* carrying the developed construct would hopefully lead to the generation of double knockout (KO) mutants for *ddm1* and *met1*, as well as single KO mutants for each gene. Though these lines were predicted to be highly sterile, as was observed in rice (Hu *et al.* 2014), they

would hopefully lead to stronger effects on meiotic recombination and would even perhaps allow for the visual of cytological meiotic phenotypes under the microscope. The generation of double mutants for *ddm1* and *met1* would also allow to determine whether a cumulative effect of both mutations would lead to stronger phenotypes than with a single mutation.

After cloning of the two sgRNAs targeting *Met1* and *Ddm1* into pBract214m (5.2.18-23), full integrity of the plasmid was validated by digesting with the *BglI* restriction enzyme (5.2.23), and the plasmid was sequenced to verify the insert was of the correct sequence. The plasmid was then transformed into *Agrobacterium tumefaciens* as described in 5.2.24 and used for embryo transformation. Unfortunately, the Genetic Transformation facility housed at the James Hutton Institute was not able to generate the CRISPR lines due to unidentified and unresolved issues at the time of writing this thesis. Therefore, the CRISPR mutants were not generated and the efficiency of the CRISPR construct could not be assessed.

5.4. Conclusion

At the start of my PhD project, there were no described mutants in DNA methylation genes *Met1* and *Ddm1* available in barley.

I used an existing TILLING population developed at the James Hutton Institute (Schreiber *et al.* 2019) to identify disrupted *met1* alleles and study their effect on DNA-methylation and meiotic recombination. The TILLING population had been developed using barley cultivar Golden Promise which was chosen as the parent cultivar because it is the reference genotype for *Agrobacterium*-mediated transformation. The idea was that any identified mutation of interest could then be transgenically complemented to confirm its functional effect. As the transformation reference, Golden Promise also has many available genomic resources, as detailed in 5.1.2, including a pseudomolecule scale genome assembly, making it an ideal candidate line for forward and reverse genetics (Schreiber *et al.* 2020).

Extensive screening of the whole TILLING population (3,072 plants) for mutations in *met1* using next generation sequencing allowed me to identify 30 lines which carried 23 unique variants in total (5.3.1.3). Five non-synonymous mutations were selected and taken forward for this project to create *met1* mutant lines. These lines were then crossed into cv. Barke, and one of these lines (P1332S) was taken further to generate F_3 families which

were used to study the effect of the mutation on meiotic recombination using genotyping with the 50K iSelect SNP array (Bayer *et al.* 2017).

As discussed in 5.3.1.5, I found crossing the lines of interest into a clean background to be difficult. The lines could have been backcrossed into the parental cultivar, Golden Promise, in order to segregate the background mutations away from the alleles of interest in each line. However, despite extensive trials, I was not successful in obtaining F_1 seeds when crossing to cv. Golden Promise, whether using the wildtype plants as female plants or pollen donor. Golden Promise is a notoriously hard variety to be used for crossing, as the spikes remain in the boot during anthesis. It is therefore necessary to dissect the spike out of stem prior to emasculating the spikelets, which often leads to the spikes drying out before pollination. These difficulties were probably heightened by the plants containing a high density of the background mutations, as well as the potential effect of the mutations of interest. This highlights one of the limitations of using Golden Promise as a cultivar for random mutagenesis: once the mutation of interest has been identified, cleaning the background backcrossing is difficult and labour intensive.

Ultimately, two mutant lines were successfully crossed into *cv.* Barke. The heterozygous F_1 Golden Promise and Barke plants facilitated recombination analysis at an earlier stage than if the mutations had been backcrossed first. However, these F_1 still carry a lot of background mutations, and there is a risk these will have segregated with the mutated allele in *met1*. Even though unlikely, this situation could lead to the observed phenotypes being caused by unidentified background mutations instead of the allele of interest.

For both of these lines, segregation of the mutation was assessed along with their potential effects on phenotypes. Neither plant height nor the number of tillers was affected by the presence of the mutations in a heterozygous state in the F_2 plants, leading to believe the mutations effect was not stringent enough to cause severe phenotypes on the plants' development. In rice, knockout mutations of *met1* leads to severe seed development impairment, large dysregulation of genome expression, and sterility (Hu *et al.* 2014). The absence of strong phenotypes in A1120T and P1332S homozygous mutant plants suggests the effect of the mutations did not lead to a knockout of the protein, providing a milder effect on the functionality of MET1. Whilst such mutations are more agronomically interesting and make the maintenance of the lines easier, the potential effect of these mutations on recombination would therefore likely be limited.

The P1332S line was used to generate F_3 families, separating the wildtype allele from the *met1* mutant allele. The F_3 families were subsequently used to determine the effect of the mutation on meiotic recombination using the 50K iSelect SNP array (Bayer *et al.* 2017) to count the number of crossover (CO) events in plants homozygous for the mutations or in wild-type plants. No significant effect of the mutation on CO number was observed, however a tendency consistent with what had previously been observed in *Arabidopsis* (Melamed-Bessudo & Levy 2012, Yelina *et al.* 2012) was noted: the number of CO events was slightly higher in sub-telomeric zone 1 and lower in pericentromeric zone 3. Such observations were also observed in VIGS knockout mutations of *met1* and *ddm1* in wheat (Raz *et al.* 2021). Confirmation of these tendencies could be obtained in the future by increasing the size of each population (wild-type and mutant) or in studying other mutations with a potentially stronger effect on the functionality of the protein.

As was discussed in Chapter 3, a method for the determination of DNA-methylation levels in plants was originally developed for this PhD thesis, but was found not to produce reliable results on *Arabidopsis* or barley DNA. It would be therefore interesting in the future to use alternative DNA-methylation measurement methods, such as Bisulfite Sequencing or Mass Spectrometry, to determine whether the identified mutations lead to a global hypomethylation of the genome in mutated plants. Similarly, a gene expression analysis of the levels of *Met1* expression in mutated plants, for example by RT-qPCR, would shed light on whether transcription of the gene has been impacted by the mutations. Similar to the microarray analysis carried out in 4.3.7, an analysis of the global gene expression profile of mutated plants would also allow to determine the extent of the effect of the mutations on plant development.

At the outset of my project, I decided to create CRISPR-Cas9 lines for both *met1* and *ddm1* which would have allowed the generation of double and single mutant lines in a clean genetic background. Combining knockout mutant alleles for both genes would likely lead to a severe reduction in DNA-methylation. On the other hand, the production of single mutants would have permitted to obtain mitigated effects of the mutations on the plants' development compared to double knockout mutants. By using a single construct to transform the lines for both genes at the same time, a significant amount of time could also have been saved, as no crossing would have been needed to generate double mutant lines after the initial transformation. CRISPR-Cas9 also presents an advantage over

random mutagenesis in that mutations are generated in a clean WT background and induced variants can be separated from the Cas9 construct by genetic segregation. Unfortunately, despite the sgRNAs and scaffold for *met1* and *ddm1* having been successfully cloned into the pBract214m CRISPR construct for barley transformation, the creation of the genetically modified lines in the FUNGEN facility failed, and no solution was found over the remainder of my PhD.

Despite the challenges that this project presented, the outcome of this research has produced several useful tools. A CRISPR-Cas9 construct was generated and cloned which would allow the generation of single and double mutant lines for *Met1* and *Ddm1*. Additionally, two TILLING mutant lines were selected and F_3 families were generated for P1332S. This line was then used to study the role of its associated mutation on the meiotic recombination landscape. Moreover, this project validated the potential of the large TILLING population developed by our research group when used to screen for mutations in a specific gene of interest, as was described in other barley cultivars (Talamè *et al.* 2008, Gottwald *et al.* 2009, Lababidi *et al.* 2009, Szarejko *et al.* 2017), and in other species such as wheat (Uauy *et al.* 2009) or *Arabidopsis* (Lai *et al.* 2012, Martín *et al.* 2009).

6. General Conclusion

6.1. Retrospective

Meiotic recombination, via the formation of crossovers (CO), plays a major role in the development of new plant varieties, and the comprehension of the mechanisms in meiosis is of paramount importance for both the breeding industry and academia. There is a large interest in understanding and manipulating the recombination landscape in plants, which could lead in turn to finely directing the introgression of traits of interest in newly developed varieties or to facilitated gene mapping in low recombination areas of the genome. The machinery directly involved in meiosis and crossover formation in plants has been thoroughly studied and, even though a lot remains to be discovered, it is now fairly well characterised, as reviewed by Mercier *et al.* (2015). However, little is known on the role of epigenetic markers, such as DNA-methylation, on the meiotic mechanisms and recombination. In *Arabidopsis*, mutants in *met1* and *ddm1* were shown to have an effect on the distribution of CO events across the genome (Melamed-Bessudo & Levy 2012, Yelina *et al.* 2012), and RNA-directed DNA-Methylation (RdDM) was shown to be dynamically regulated during gametogenesis (Walker *et al.* 2018). Nevertheless, at the start of this PhD, no studies had been conducted on the effect of DNA-methylation on meiosis in complex genome crops such as wheat, rice, or barley.

Barley, compared to wheat for example, is a cereal with a relatively small, simple, diploid genome (5.3 Gb, 2n = 2x =14). This makes it an ideal model for the study of meiotic recombination in cereals, and several studies on barley meiosis have been conducted over the years (Phillips *et al.* 2012, Higgins 2013, Higgins *et al.* 2012, Higgins *et al.* 2014, Bennett *et al.* 1973, Phillips *et al.* 2015, Colas *et al.* 2016, Barakate *et al.* 2021, Colas *et al.* 2019, Mittmann 2017, Phillips *et al.* 2013, Tsubouchi *et al.* 2006, Barakate *et al.* 2014, Colas *et al.* 2017). Moreover, during this PhD project, several resources were produced to facilitate the use of barley as a research organism, including a 50K iSelect SNP genotyping chip (Bayer *et al.* 2017) and a new genome assembly for "transformation reference" *cv*. Golden Promise (Schreiber *et al.* 2020). The tools were paramount in the progress of this thesis.

The overall goal of this thesis was to elucidate the role of DNA-methylation on the meiotic recombination landscape in barley (*Hordeum vulgare*) and *Arabidopsis*. In a first instance,

zebularine, a chemical inducing global demethylation of the genome, was applied to *Arabidopsis* and barley seeds during germination. Plants exposed to zebularine transiently phenocopied *Arabidopsis* and rice *met1* mutants (Mirouze *et al.* 2012, Hu *et al.* 2014) in that they exhibited delayed development and seed lethality. However, both *Arabidopsis* and barley plants recovered a normal phenotype when the zebularine treatment was interrupted. In both species, gene expression patterns were disrupted in seedlings which had been exposed to various concentrations of zebularine compared to untreated plantlets. In *Arabidopsis*, recombination rates were decreased in sub-telomeric regions, but undisturbed in interstitial and centromeric regions, whereas CO frequency appeared unaffected in zebularine-treated barley F_1 hybrids.

Secondly, a library of *met1* mutants in barley was generated by screening for variants in the *Met1* gene across a large TILLING population and wild barley accessions. Golden Promise TILLING mutant lines A1120T and P1332S were selected to cross into *cv*. Barke in order to study the effect of the mutations on plant phenotypes and meiotic recombination. Although plant height and the number of tillers was not affected by the presence of the *met1* mutation in either of the lines, analysis of the recombination landscape in P1332S F_3 families showed a non-significant tendency to shifted CO patterns towards sub-telomeric regions, with slightly decreased levels of recombination in centromeric regions.

In parallel, a range of different tools and techniques were developed and optimised during this PhD thesis, which will facilitate future studies of the role of DNA-methylation on meiotic recombination in plants. A new fluorescent-based restriction assay for DNA-methylation was adapted from use on mammalian cells (Zhou *et al.* 2017) to be applied to plant genomic DNA. This protocol still requires optimisation, and perhaps more reliable equipment, however once adjusted this assay would allow for a cost-efficient routine measurement of global DNA-methylation levels in hypomethylated plants, either chemically treated or mutagenized. This project also confirmed the benefit of chemically inducing hypomethylation in plants with zebularine, which affects the plants in a transient manner. When fine-tuning the methods and time of delivery of the chemical to the plants, chemicals such as zebularine and its analogue 5-azacytidine could be used to transiently modify DNA-methylation levels at key developmental stages such as meiosis, allowing for recovery and hopefully, easier maintenance of the subsequent generation

when compared to hypomethylated mutants. Finally, in this PhD project I identified a large collection of induced and natural variants in *met1*, through screening a large TILLING population and a diverse collection of barley accessions. This highlighted two main techniques which can be used to identify mutant plants in reverse genetics approaches: single-gene sequencing and Exome Capture. In particular, this thesis project allowed for the optimisation of the full sequencing of a single gene across 768 plants in a single sequencing reaction, through multiplexing and a strong pooling/data deconvolution approach. This highlighted the possibility of effectively screening for variants within a large population in a time and cost-efficient manner. This work resulted in a peer-reviewed publication of this TILLING population (Schreiber *et al.* 2019).

6.2. Perspective

Whilst the work carried out during this PhD seems to confirm a relationship between DNA-methylation, plant development, and meiotic recombination, a lot remains uncovered and would necessitate further investigation.

The first logical step would be to fully characterise the extent to which zebularine treatment and the identified mutations impact DNA-methylation levels in plants. As explained in Chapter 3, a new protocol for the measurement of DNA-methylation levels has been adapted to plant material, however it still requires optimisation. It would be interesting to assess whether different fluorescent probes, for example Alexa-647-dCTPs or Alexa-555-dCTPs, which were used in the original study (Zhou *et al.* 2017), would prove to be more stable during the fluorescence reading stage. On the other hand, other plate readers could also be used to collect fluorescence levels in the samples, in order to verify whether the lack of sensitivity of the assay is due to technical limitations. Finally, it has not been determined yet whether the assay was sensitive enough to detect hypomethylation in plant genomes, and to which extent it would correlate with pre-existing data such as bisulfite-sequencing data. Therefore, the assay should be conducted on already characterised hypomethylated DNA, like that from *Arabidopsis met1-3* mutants. Seeds from this population had been obtained from Prof. Ian Henderson (University of Cambridge) in the final year of this PhD project. However, due to time constraints, they could not be used to validate this fluorescent-based DNA-methylation assay.

The delivery of zebularine to germinating seeds in *Arabidopsis* and barley was shown to phenocopy rice and *Arabidopsis met1* mutants in that there were high levels of lethality, the development of the seeds was impaired, and the plants looked sickly compared to untreated seeds (Mirouze *et al.* 2009, Hu *et al.* 2014). However, as had been previously described in zebularine-treated seedlings (Baubec *et al.* 2009), the plants recovered normal phenotypes after interruption of the chemical treatment, for both barley and *Arabidopsis* plants. This resulted in very mild or inexistent effects of the treatment on meiotic recombination in both species. One way to address this issue would be to precisely target meiotic stages when treating the plants with zebularine. A "Needle and Thread" technique was tested on barley F1 hybrids "Jettoo" on the advice of Dr. Stefan Heckmann (IPK Gatersleben) in order to continuously supply the developing spikes with zebularine during the meiotic stages. This method proved successful in delivering 5-ethynyl-2'-deoxyuridine (EdU) in barley meiocytes (Ahn *et al.* 2020). However, it would have required further optimisation to adapt this technique to winter F_1 hybrids, for which identifying the ideal meiotic stages proved difficult without dissecting the stem to check for correct delivery of the chemical.

As highlighted in Chapter 5, several mutant lines for *met1* were identified by screening for variants in a large TILLING population as well as in a collection of wild barley accessions. All these lines would require further analysis to determine to what extent the identified mutations cause an effect on DNA-methylation in the plants. In the first instance, the mutants identified from the TILLING population would need to be backcrossed in a clean parental background (*cv*. Golden Promise) in an attempt to segregate the mutation of interest away from other background mutations caused by EMS mutagenesis. It would be interesting to analyse gene expression patterns in these mutant lines, for *Met1* in particular, but also for other genes previously identified as being affected by DNA-methylation levels (such as those identified in 4.3.7). This would allow to determine whether a correlation can be observed between the type of mutation obtained and the effect on both the transcription levels of the protein itself, and gene expression regulation across the genome. Future work should also include analysis of DNA-methylation levels in these mutant lines, either using a further optimised version of the protocol discussed in Chapter 3 or by other techniques such as bisulfite sequencing or mass spectrometry. For lines where a strong correlation would be found between the identified mutation and hypomethylation phenotypes, the mutated and wild-type

proteins could be expressed *in vitro* to allow for the study of the impact of the mutation on the protein's structure and functionality. For example, an enzymatic methyltransferase activity assay such as Methyltransferase-GLO (Hsiao *et al.* 2016) on purified mutant and wild-type proteins would allow to determine whether the mutation impacts the enzymatic activity of the *Met1* protein itself.

Genetic transformation of barley plants using CRISPR-Cas9 vectors targeting both *met1* and *ddm1* was initiated during this PhD project (5.3.5). However, due to technical difficulties, the efficiency of the obtained CRISPR-Cas9 construct could not be assessed, and no mutant plants were obtained. In the future, this construct could be used to generate double and single mutants for *met1* and *ddm1* in barley, in order to assess the role of these two genes, separately or combined, on plant development and meiotic recombination, in barley. Other genetic transformation tools could also be used to study the role of these genes in barley and *Arabidopsis*. Due to its higher susceptibility to transformation, *Arabidopsis* could be used to overexpress *Met1*, either constitutively or transiently, in order to study the role of hypermethylation in plants on development and meiosis. Traditional *Arabidopsis* transformation has previously been used to constitutively overexpress *Met1*, inducing abnormal phenotypes and strong changes in gene expression levels (Brocklehurst *et al.* 2018). However more work could be done to determine the role of *Met1* on meiotic recombination in these overexpression lines, or by using a meiocyte-specific promoter to guide *Met1* expression in *Arabidopsis* plants. For example, the meiosis-specific DMC1 promoter was used to generate CRISPR-Cas9 mutants in *tb7*, *zyp1* and *smc3* in Maize (Feng *et al.* 2018). Finally other gene editing techniques could be used to silence methylation genes *met1* and *ddm1* in barley plants, such as delivery of hairpins targeting each gene through RNAi lines or VIGS. At the time of the submission of this thesis, a recent study showed efficient transient silencing of *met1* and *ddm1* in wheat (*Triticum aestivum*) led to increased recombination frequency in the sub-telomeric region of the short arm of chromosome 1A (Raz *et al.* 2021). Such techniques could be transferred to barley transformation reference cultivar Golden Promise, in order to transiently affect DNA-methylation and assess its effect on meiotic recombination. This would be a way to remobilise the plasticity of the barley genome to develop new cultivars that are more productive in a more sustainable agriculture context.

7. References

AGARWAL, S. & ROEDER, G.S. 2000. Zip3 provides a link between recombination enzymes and synaptonemal complex proteins. *Cell*, **102**, 245–255, 10.1016/S0092-8674(00)00029-5.

AGRAWAL, N., DASARADHI, P.V.N., MOHMMED, A., MALHOTRA, P., BHATNAGAR, R.K. & MUKHERJEE, S.K. 2003. RNA Interference: Biology, Mechanism, and Applications. *Microbiology and Molecular Biology Reviews*, **67**, 657–685, 10.1128/mmbr.67.4.657-685.2003.

AHN, Y.J., CUACOS, M., AYOUB, M.A., KAPPERMANN, J., HOUBEN, A. & HECKMANN, S. 2020. In Planta Delivery of Chemical Compounds into Barley Meiocytes: EdU as Compound Example. *In Methods in Molecular Biology*. Humana Press Inc., 381–402., 10.1007/978-1-4939-9818-0_27.

AKEY, D.T., AKEY, J.M., ZHANG, K. & JIN, L. 2002. Assaying DNA methylation based on high-throughput melting curve approaches. *Genomics*, **80**, 376–384, 10.1006/geno.2002.6851.

ALLERS, T. & LICHTEN, M. 2001. Differential timing and control of noncrossover and crossover recombination during meiosis. *Cell*, **106**, 47–57, 10.1016/S0092-8674(01)00416-0.

APPELS, R., EVERSOLE, K., FEUILLET, C., KELLER, B., ROGERS, J., STEIN, N., POZNIAK, C.J., et al. 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, **361**, 10.1126/science.aar7191.

ARMSTRONG, S.J. & JONES, G.H. 2001. Female meiosis in wild-type Arabidopsis thaliana and in two meiotic mutants. *Sexual Plant Reproduction*, **13**, 177–183, 10.1007/s004970000050.

AUSIN, I., MOCKLER, T.C., CHORY, J. & JACOBSEN, S.E. 2009. IDN1 and IDN2 are required for De Novo DNA methylation in Arabidopsis Thaliana. *Nature Structural and Molecular Biology*, **16**, 1325–1327, 10.1038/nsmb.1690.

BAIK, B.K. & ULLRICH, S.E. 2008. Barley for food: Characteristics, improvement, and renewed interest. *Journal of Cereal Science*, **48**, 233–242, 10.1016/j.jcs.2008.02.002.

BAKER, K., BAYER, M., COOK, N., DREIßIG, S., DHILLON, T., RUSSELL, J., HEDLEY, P.E., et al. 2014. The low-recombining pericentromeric region of barley restricts gene diversity and evolution but not gene expression. *Plant Journal*, **79**, 981–992, 10.1111/tpj.12600.

BARAKATE, A., HIGGINS, J.D., VIVERA, S., STEPHENS, J., PERRY, R.M., RAMSAY, L., COLAS, I., et al. 2014. The synaptonemal complex protein ZYP1 is required for imposition of meiotic crossovers in barley. *Plant Cell*, **26**, 729–740, 10.1105/tpc.113.121269.

BARAKATE, A., ORR, J., SCHREIBER, M., COLAS, I., LEWANDOWSKA, D., MCCALLUM, N., MACAULAY, M., et al. 2021. Barley Anther and Meiocyte Transcriptome Dynamics in Meiotic Prophase I. *Frontiers in Plant Science*, **11**, 619404, 10.3389/fpls.2020.619404.

BASS, H.W. 2003. Telomere dynamics unique to meiotic prophase: Formation and significance of the bouquet. *Cellular and Molecular Life Sciences*, **60**, 2319–2324, 10.1007/s00018-003-3312-4.

BAUBEC, T., FINKE, A., MITTELSTEN SCHEID, O. & PECINKA, A. 2014. Meristem-specific expression of epigenetic regulators safeguards transposon silencing in Arabidopsis. *EMBO Reports*, **15**, 446–452, 10.1002/embr.201337915.

BAUBEC, T., PECINKA, A., ROZHON, W. & MITTELSTEN SCHEID, O. 2009. Effective, homogeneous and transient interference with cytosine methylation in plant genomic DNA by zebularine. *Plant Journal*, **57**, 542–554, 10.1111/j.1365-313X.2008.03699.x.

BAYER, M.M., RAPAZOTE-FLORES, P., GANAL, M., HEDLEY, P.E., MACAULAY, M., PLIESKE, J., RAMSAY, L., et al. 2017. Development and evaluation of a barley 50k iSelect SNP array. *Frontiers in Plant Science*, **8**, 1792, 10.3389/fpls.2017.01792.

BEIER, S., HIMMELBACH, A., COLMSEE, C., ZHANG, X.Q., BARRERO, R.A., ZHANG, Q., LI, L., et al. 2017. Construction of a map-based reference genome sequence for barley, Hordeum vulgare L. *Scientific Data*, **4**, 1–24, 10.1038/sdata.2017.44.

BENNETT, M.D., FINCH, R.A., SMITH, J.B. & RAO, M.K. 1973. The time and duration of female meiosis in wheat, rye and barley. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, **183**, 301–319, 10.1098/rspb.1973.0019.

BERANEK, D.T., WEIS, C.C. & SWENSON, D.H. 1980. A comprehensive quantitative analysis of methylated and ethylated DNA using high pressure liquid chromatography. *Carcinogenesis*, **1**, 595–606, 10.1093/carcin/1.7.595.

BHUIYAN, H. & SCHMEKEL, K. 2004. Meiotic chromosome synapsis in yeast can occur without Spo11-induced DNA double-strand breaks. *Genetics*, **168**, 775–783, 10.1534/genetics.104.029660.

BIES-ETHEVE, N., PONTIER, D., LAHMY, S., PICART, C., VEGA, D., COOKE, R. & LAGRANGE, T. 2009. RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Reports*, **10**, 649–654, 10.1038/embor.2009.31.

BLUM, M., CHANG, H.Y., CHUGURANSKY, S., GREGO, T., KANDASAAMY, S., MITCHELL, A., NUKA, G., et al. 2021. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, **49**, D344–D354, 10.1093/nar/gkaa977.

BOLGER, A.M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120, 10.1093/bioinformatics/btu170.

BOOTH, A. 2016. Barley Crossing - YouTube Available at: https://www.youtube.com/watch?v=HQ_-0ALP1ig&ab_channel=BarleyHub [Accessed April 17, 2021].

BOUMA, J. 1967. New Variety of Spring Barley Diamant in Czechoslovakia. *Abhandlungen der Deutschen Akademie der Wissenschaften zu Berlin*, **2**, 177–182.

BRADSHAW, J.E. 2017. Plant breeding: past, present and future. *Euphytica*, **213**, 10.1007/s10681-016-1815-y.

BROCKLEHURST, S., WATSON, M., CARR, I.M., OUT, S., HEIDMANN, I. & MEYER, P. 2018. Induction of epigenetic variation in Arabidopsis by over-expression of DNA METHYLTRANSFERASE1 (MET1). *PLoS ONE*, **13**, 10.1371/journal.pone.0192170.

BROWN, M.S., GRUBB, J., ZHANG, A., RUST, M.J. & BISHOP, D.K. 2015. Small Rad51 and Dmc1 Complexes Often Co-occupy Both Ends of a Meiotic DNA Double Strand Break Lichten, M., *ed. PLoS Genetics*, **11**, e1005653, 10.1371/journal.pgen.1005653.

CALDWELL, D.G., MCCALLUM, N., SHAW, P., MUEHLBAUER, G.J., MARSHALL, D.F. & WAUGH, R. 2004. A structured mutant population for forward and reverse genetics in Barley (Hordeum vulgare L.). *Plant Journal*, **40**, 143–150, 10.1111/j.1365-313X.2004.02190.x.

CALLEBAUT, I., COURVALIN, J.C. & MORNON, J.P. 1999. The BAH (bromo-adjacent homology) domain: A link between DNA methylation, replication and transcriptional regulation.

FEBS Letters, **446**, 189–193, 10.1016/S0014-5793(99)00132-5.

CHAMBERS, A.L., PEARL, L.H., OLIVER, A.W. & DOWNS, J.A. 2013. The BAH domain of Rsc2 is a histone H3 binding domain. *Nucleic Acids Research*, **41**, 9168–9182, 10.1093/nar/gkt662.

CHAMPION, C., GUIANVARC'H, D., SÉNAMAUD-BEAUFORT, C., JURKOWSKA, R.Z., JELTSCH, A., PONGER, L., ARIMONDO, P.B. & GUIEYSSE-PEUGEOT, A.L. 2010. Mechanistic insights on the inhibition of C5 DNA methyltransferases by zebularine Mayer, C., *ed. PLoS ONE*, **5**, 12388, 10.1371/journal.pone.0012388.

CHARI, R., YEO, N.C., CHAVEZ, A. & CHURCH, G.M. 2017. SgRNA Scorer 2.0: A Species-Independent Model to Predict CRISPR/Cas9 Activity. *ACS Synthetic Biology*, **6**, 902–904, 10.1021/acssynbio.6b00343.

CHEN, L., MACMILLAN, A.M., CHANG, W., EZAZ-NIKPAY, K., VERDINE, G.L. & LANE, W.S. 1991. Direct Identification of the Active-Site Nucleophile in a DNA (Cytosine-5)-methyltransferase. *Biochemistry*, **30**, 11018–11025, 10.1021/bi00110a002.

CHENG, J.C., YOO, C.B., WEISENBERGER, D.J., CHUANG, J., WOZNIAK, C., LIANG, G., MARQUEZ, V.E., et al. 2004. Preferential response of cancer cells to zebularine. *Cancer Cell*, **6**, 151–158, 10.1016/j.ccr.2004.06.023.

CHO, S.-W., ISHII, T., MATSUMOTO, N., TANAKA, H., ELTAYEB, A. & TSUJIMOTO, H. 2012. Effects of the cytidine analogue zebularine on wheat mitotic chromosomes. *Chromosome Science*, **14**, 23–28, 10.11352/scr.14.23.

CHOI, K., ZHAO, X., TOCK, A.J., LAMBING, C., UNDERWOOD, C.J., HARDCASTLE, T.J., SERRA, H., et al. 2018. Nucleosomes and DNA methylation shape meiotic DSB frequency in Arabidopsis thaliana transposons and gene regulatory regions. *Genome Research*, **28**, 532–546, 10.1101/gr.225599.117.

CHOI, Y. & CHAN, A.P. 2015. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, **31**, 2745–2747, 10.1093/bioinformatics/btv195.

CHOI, Y., SIMS, G.E., MURPHY, S., MILLER, J.R. & CHAN, A.P. 2012. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE*, **7**, 10.1371/journal.pone.0046688.

CHOULET, F., ALBERTI, A., THEIL, S., GLOVER, N., BARBE, V., DARON, J., PINGAULT, L., et al. 2014. Structural and functional partitioning of bread wheat chromosome 3B. *Science*, **345**, 10.1126/science.1249721.

COKUS, S.J., FENG, S., ZHANG, X., CHEN, Z., MERRIMAN, B., HAUDENSCHILD, C.D., PRADHAN, S., NELSON, S.F., PELLEGRINI, M. & JACOBSEN, S.E. 2008. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, **452**, 215–219, 10.1038/nature06745.

COLAS, I., BARAKATE, A., MACAULAY, M., SCHREIBER, M., STEPHENS, J., VIVERA, S., HALPIN, C., WAUGH, R. & RAMSAY, L. 2019. desynaptic5 carries a spontaneous semi-dominant mutation affecting Disrupted Meiotic cDNA 1 in barley. *Journal of Experimental Botany*, **70**, 2683–2698, 10.1093/jxb/erz080.

COLAS, I., DARRIER, B., ARRIETA, M., MITTMANN, S.U., RAMSAY, L., SOURDILLE, P. & WAUGH, R. 2017. Observation of extensive chromosome axis remodeling during the "Diffuse-phase" of Meiosis in large genome cereals. *Frontiers in Plant Science*, **8**, 1235, 10.3389/fpls.2017.01235.

COLAS, I., MACAULAY, M., HIGGINS, J.D., PHILLIPS, D., BARAKATE, A., POSCH, M., ARMSTRONG, S.J., et al. 2016. A spontaneous mutation in MutL-Homolog 3 (HvMLH3) affects synapsis and crossover resolution in the barley desynaptic mutant des10. *New Phytologist*, **212**, 693–707, 10.1111/nph.14061.

COLAS, I., SHAW, P., PRIETO, P., WANOUS, M., SPIELMEYER, W., MAGO, R. & MOORE, G. 2008. Effective chromosome pairing requires chromatin remodeling at the onset of meiosis. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 6075–6080, 10.1073/pnas.0801521105.

COLMSEE, C., BEIER, S., HIMMELBACH, A., SCHMUTZER, T., STEIN, N., SCHOLZ, U. & MASCHER, M. 2015. BARLEX - The barley draft genome explorer. *Molecular Plant*, **8**, 964–966, 10.1016/j.molp.2015.03.009.

COLOME-TATCHE, M., CORTIJO, S., WARDENAAR, R., MORGADO, L., LAHOUZ, B., SARAZIN, A., ETCHEVERRY, M., et al. 2012. Features of the Arabidopsis recombination landscape resulting from the combined loss of sequence variation and DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 16240–16245, 10.1073/pnas.1212955109.

COMADRAN, J., KILIAN, B., RUSSELL, J., RAMSAY, L., STEIN, N., GANAL, M., SHAW, P., et al. 2012. Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. *Nature Genetics*, **44**, 1388–1391, 10.1038/ng.2447.

COMADRAN, J., RAMSAY, L., MACKENZIE, K., HAYES, P., CLOSE, T.J., MUEHLBAUER, G., STEIN, N. & WAUGH, R. 2011. Patterns of polymorphism and linkage disequilibrium in cultivated barley. *Theoretical and Applied Genetics*, **122**, 523–531, 10.1007/s00122-010-1466-7.

COREM, S., DORON-FAIGENBOIM, A., JOUFFROY, O., MAUMUS, F., ARAZI, T. & BOUCHE, N. 2018. Redistribution of CHH methylation and small interfering RNAs across the genome of tomato ddm1 mutants. *Plant Cell*, **30**, 1628–1644, 10.1105/tpc.18.00167.

CORPET, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research*, **16**, 10881–10890, 10.1093/nar/16.22.10881.

DA INES, O., ABE, K., GOUBELY, C., GALLEGO, M.E. & WHITE, C.I. 2012. Differing requirements for RAD51 and DMC1 in meiotic pairing of centromeres and chromosome arms in Arabidopsis thaliana Pawlowski, W.P., *ed. PLoS Genetics*, **8**, e1002636, 10.1371/journal.pgen.1002636.

DALY, P., MCCLELLAN, C., MALUK, M., OAKEY, H., LAPIERRE, C., WAUGH, R., STEPHENS, J., et al. 2019. RNAi-suppression of barley caffeic acid O-methyltransferase modifies lignin despite redundancy in the gene family. *Plant Biotechnology Journal*, **17**, 594–607, 10.1111/pbi.13001.

DANECEK, P., BONFIELD, J.K., LIDDLE, J., MARSHALL, J., OHAN, V., POLLARD, M.O., WHITWHAM, A., et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience*, **10**, 1–4, 10.1093/gigascience/giab008.

DE MASSY, B. 2013. Initiation of meiotic recombination: How and where? Conservation and specificities among eukaryotes. *Annual Review of Genetics*, **47**, 563–599, 10.1146/annurev-genet-110711-155423.

DRUKA, A., FRANCKOWIAK, J., LUNDQVIST, U., BONAR, N., ALEXANDER, J., HOUSTON, K., RADOVIC, S., et al. 2011. Genetic dissection of barley morphology and development. *Plant Physiology*, **155**, 617–627, 10.1104/pp.110.166249.

DU, J., ZHONG, X., BERNATAVICHUTE, Y. V., STROUD, H., FENG, S., CARO, E., VASHISHT, A.A., et al.

2012. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell*, **151**, 167–180, 10.1016/j.cell.2012.07.034.

DU, M., LUO, M., ZHANG, R., FINNEGAN, E.J. & KOLTUNOW, A.M.G. 2014. Imprinting in rice: The role of DNA and histone methylation in modulating parent-of-origin specific expression and determining transcript start sites. *Plant Journal*, **79**, 232–242, 10.1111/tpj.12553.

DUBROVINA, A.S. & KISELEV, K. V. 2019. Exogenous RNAs for gene regulation and plant resistance. *International Journal of Molecular Sciences*, **20**, 10.3390/ijms20092282.

DÜRR, H., FLAUS, A., OWEN-HUGHES, T. & HOPFNER, K.P. 2006. Snf2 family ATPases and DExx box helicases: Differences and unifying concepts from high-resolution crystal structures. *Nucleic Acids Research*, **34**, 4160–4167, 10.1093/nar/gkl540.

EICHTEN, S.R., BRISKINE, R., SONG, J., LI, Q., SWANSON-WAGNER, R., HERMANSON, P.J., WATERS, A.J., et al. 2013. Epigenetic and Genetic Influences on DNA Methylation Variation in Maize Populations. *Plant Cell*, **25**, 2783–2797, 10.1105/tpc.113.114793.

EISEN, J.A., SWEDER, K.S. & HANAWALT, P.C. 1995. Evolution of the SNF2 family of proteins: Subfamilies with distinct sequences and functions. *Nucleic Acids Research*, **23**, 2715–2723, 10.1093/nar/23.14.2715.

ERHARD, K.F., TALBOT, J.E.R.B., DEANS, N.C., MCCLISH, A.E. & HOLLICK, J.B. 2015. Nascent transcription affected by RNA polymerase IV in Zea mays. *Genetics*, **199**, 1107–1125, 10.1534/genetics.115.174714.

FAO. 2020. FAOSTAT Available at: http://www.fao.org/faostat/en/#data/QC/visualize [Accessed December 21, 2016].

FENG, C., SU, H., BAI, H., WANG, R., LIU, Y., GUO, X., LIU, C., et al. 2018. High-efficiency genome editing using a *dmc1* promoter-controlled CRISPR/Cas9 system in maize. *Plant Biotechnology Journal*, **16**, 1848–1857, 10.1111/pbi.12920.

FENG, S., COKUS, S.J., SCHUBERT, V., ZHAI, J., PELLEGRINI, M. & JACOBSEN, S.E. 2014. Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis. *Molecular Cell*, **55**, 694–707, 10.1016/j.molcel.2014.07.008.

FINNEGAN, E.J., FORD, B., WALLACE, X., PETTOLINO, F., GRIFFIN, P.T., SCHMITZ, R.J., ZHANG, P., et al. 2018. Zebularine treatment is associated with deletion of FT-B1 leading to an increase in spikelet number in bread wheat. *Plant Cell and Environment*, **41**, 1346–1360, 10.1111/pce.13164.

FLOTHO, C., CLAUS, R., BATZ, C., SCHNEIDER, M., SANDROCK, I., IHDE, S., PLASS, C., NIEMEYER, C.M. & LÜBBERT, M. 2009. The DNA methyltransferase inhibitors azacitidine, decitabine and zebularine exert differential effects on cancer gene expression in acute myeloid leukemia cells. *Leukemia*, **23**, 1019–1028, 10.1038/leu.2008.397.

FORSTER, B.P. 2001. Mutation genetics of salt tolerance in barley: An assessment of Golden Promise and other semi-dwarf mutants. *Euphytica*, **120**, 317–328, 10.1023/A:1017592618298.

FOSTER, A.E. 1987. Effects of a semidwarf gene from Jotun on agronomic and quality traits of barley. *In Proceedings of the 5^< th> International Barley Genetics Symposium, 1986, Sanyo, Okayama*. 979–982.

GALE, M.D. & REES, H. 1970. Genes controlling chiasma frequency in hordeum. *Heredity*, **25**, 393–410, 10.1038/hdy.1970.40.

GALLEGO-BARTOLOMÉ, J., GARDINER, J., LIU, W., PAPIKIAN, A., GHOSHAL, B., KUO, H.Y., ZHAO, J.M.C., SEGAL, D.J. & JACOBSEN, S.E. 2018. Targeted DNA demethylation of the arabidopsis genome using the human TET1 catalytic domain. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, E2125–E2134, 10.1073/pnas.1716945115.

GALLEGO-BARTOLOMÉ, J. 2020. DNA methylation in plants: mechanisms and tools for targeted manipulation. *New Phytologist*, **227**, 38–44, 10.1111/nph.16529.

GALLEGO, M.E., JEANNEAU, M., GRANIER, F., BOUCHEZ, D., BECHTOLD, N. & WHITE, I. 2001. Disruption of the Arabidopsis RAD50 gene leads to plant sterility and MMS sensitivity . *The Plant Journal*, **25**, 31–41, 10.1111/j.1365-313x.2001.00928.x.

GAO, G., BI, X., CHEN, J., SRIKANTA, D. & RONG, Y.S. 2009. Mre11-Rad50-Nbs complex is required to cap telomeres during Drosophila embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 10728–10733, 10.1073/pnas.0902707106.

GAO, Z., LIU, H.L., DAXINGER, L., PONTES, O., HE, X., QIAN, W., LIN, H., et al. 2010. An RNA

polymerase II-and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature*, **465**, 106–109, 10.1038/nature09025.

GARCIA-GIMENEZ, G., BARAKATE, A., SMITH, P., STEPHENS, J., KHOR, S.F., DOBLIN, M.S., HAO, P., et al. 2020. Targeted mutation of barley (1,3;1,4)-β-glucan synthases reveals complex relationships between the storage and cell wall polysaccharide content. *The Plant Journal*, **104**, 1009–1022, 10.1111/tpj.14977.

GARRISON, E. & MARTH, G. 2012. Haplotype-based variant detection from short-read sequencing, 1–9 pp. Available at: http://arxiv.org/abs/1207.3907.

GEHRING, M., BUBB, K.L. & HENIKOFF, S. 2009. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science*, **324**, 1447–1451, 10.1126/science.1171609.

GNYSZKA, A., JASTRZĘBSKI, Z. & FLIS, S. 2013. DNA methyltransferase inhibitors and their emerging role in epigenetic therapy of cancer. *Anticancer Research*, **33**, 2989–2996 Available at: http://www.ncbi.nlm.nih.gov/pubmed/23898051 [Accessed October 19, 2018].

GOL, L., HARALDSSON, E.B. & VON KORFF, M. 2021. *Ppd-H1* integrates drought stress signals to control spike development and flowering time in barley Jones, M., *ed. Journal of Experimental Botany*, **72**, 122–136, 10.1093/jxb/eraa261.

GONG, Z., MORALES-RUIZ, T., ARIZA, R.R., ROLDÁN-ARJONA, T., DAVID, L. & ZHU, J.K. 2002. ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. *Cell*, **111**, 803–814, 10.1016/S0092-8674(02)01133-9.

GOTTWALD, S., BAUER, P., KOMATSUDA, T., LUNDQVIST, U. & STEIN, N. 2009. TILLING in the two-rowed barley cultivar "Barke" reveals preferred sites of functional diversity in the gene HvHox1. *BMC Research Notes*, **2**, 258, 10.1186/1756-0500-2-258.

GRIFFIN, P.T., NIEDERHUTH, C.E. & SCHMITZ, R.J. 2016. A comparative analysis of 5-azacytidine-and zebularine-induced DNA demethylation. *G3: Genes, Genomes, Genetics*, **6**, 2773–2780, 10.1534/g3.116.030262.

GROB, S., SCHMID, M.W. & GROSSNIKLAUS, U. 2014. Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila. *Molecular Cell*, **55**, 678–693, 10.1016/j.molcel.2014.07.009.

GROTH, M., MOISSIARD, G., WIRTZ, M., WANG, H., GARCIA-SALINAS, C., RAMOS-PARRA, P.A., BISCHOF, S., et al. 2016. MTHFD1 controls DNA methylation in Arabidopsis. *Nature Communications*, **7**, 1–13, 10.1038/ncomms11640.

GUNUPURU, L.R., PEROCHON, A., ALI, S.S., SCOFIELD, S.R. & DOOHAN, F.M. 2019. Virus-induced gene silencing (VIGS) for functional characterization of disease resistance genes in barley seedlings. *In Methods in Molecular Biology*. Humana Press Inc., 95–114., 10.1007/978-1-4939-8944-7_7.

GUO, W., TZIOUTZIOU, N.A., STEPHEN, G., MILNE, I., CALIXTO, C.P.G., WAUGH, R., BROWN, J.W.S. & ZHANG, R. 2020. 3D RNA-seq: a powerful and flexible tool for rapid and accurate differential expression and alternative splicing analysis of RNA-seq data for biologists. *RNA Biology*, 10.1101/656686.

HAAHR, V. 1976. Studies of an induced, high-yielding dwarf-mutant of spring barley. *In Proc. Int. Barley Genet. Symp., 3^< rd> Garching, Federal Republic of Germany, 7-12 July 1975*. Verlag Karl Thieming.

HARWOOD, W.A. 2014. A protocol for high-throughput agrobacterium-mediated barley transformation. *Methods in Molecular Biology*, **1099**, 251–260, 10.1007/978-1-62703-715-0_20.

HERMAN, J.G., GRAFF, J.R., MYÖHÄNEN, S., NELKIN, B.D. & BAYLIN, S.B. 1996. Methylation-specific PCR: A novel PCR assay for methylation status of CpG islands. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 9821–9826, 10.1073/pnas.93.18.9821.

HIGGINS, J.D. 2013. Analyzing meiosis in barley. *Methods in Molecular Biology*, **990**, 135–144, 10.1007/978-1-62703-333-6_14.

HIGGINS, J.D., ARMSTRONG, S.J., FRANKLIN, F.C.H. & JONES, G.H. 2004. The Arabidopsis MutS homolog AtMSH4 functions at an early step in recombination: Evidence for two classes of recombination in Arabidopsis. *Genes and Development*, **18**, 2557–2570, 10.1101/gad.317504.

HIGGINS, J.D., OSMAN, K., JONES, G.H. & FRANKLIN, F.C.H. 2014. Factors underlying restricted crossover localization in barley meiosis. *Annual Review of Genetics*, **48**, 29–47, 10.1146/annurev-genet-120213-092509.

HIGGINS, J.D., PERRY, R.M., BARAKATE, A., RAMSAY, L., WAUGH, R., HALPIN, C., ARMSTRONG, S.J. &

FRANKLIN, F.C.H. 2012. Spatiotemporal asymmetry of the meiotic program underlies the predominantly distal distribution of meiotic crossovers in barley. *Plant Cell*, **24**, 4096–4109, 10.1105/tpc.112.102483.

HISANO, H. & SATO, K. 2016. Genomic regions responsible for amenability to Agrobacterium-mediated transformation in barley. *Scientific Reports*, **6**, 1–11, 10.1038/srep37505.

HOLME, I.B., WENDT, T., GIL-HUMANES, J., DELEURAN, L.C., STARKER, C.G., VOYTAS, D.F. & BRINCH-PEDERSEN, H. 2017. Evaluation of the mature grain phytase candidate HvPAPhy_a gene in barley (Hordeum vulgare L.) using CRISPR/Cas9 and TALENs. *Plant Molecular Biology*, **95**, 111–121, 10.1007/s11103-017-0640-6.

HSIAO, K., ZEGZOUTI, H. & GOUELI, S.A. 2016. Methyltransferase-Glo: a universal, bioluminescent and homogenous assay for monitoring all classes of methyltransferases. *Epigenomics*, **8**, 321–339, 10.2217/epi.15.113.

HSIEH, P.H., HE, S., BUTTRESS, T., GAO, H., COUCHMAN, M., FISCHER, R.L., ZILBERMAN, D. & FENG, X. 2016. Arabidopsis male sexual lineage exhibits more robust maintenance of CG methylation than somatic tissues. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 15132–15137, 10.1073/pnas.1619074114.

HSIEH, T.F., IBARRA, C.A., SILVA, P., ZEMACH, A., ESHED-WILLIAMS, L., FISCHER, R.L. & ZILBERMAN, D. 2009. Genome-wide demethylation of Arabidopsis endosperm. *Science*, **324**, 1451–1454, 10.1126/science.1172417.

HU, L., LI, N., XU, C., ZHONG, S., LIN, X., YANG, J., ZHOU, T., et al. 2014. Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. *Proceedings of the National Academy of Sciences*, **111**, 10642–10647, 10.1073/pnas.1410761111.

HUANG, B., HUANG, D., HONG, Z., OWIE, S.O. & WU, W. 2020. Genetic analysis reveals four interacting loci underlying awn trait diversity in barley (Hordeum vulgare). *Scientific Reports*, **10**, 12535, 10.1038/s41598-020-69335-x.

HYE, R.W., PONTES, O., PIKAARD, C.S. & RICHARDS, E.J. 2007. VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. *Genes and Development*, **21**, 267–277, 10.1101/gad.1512007.

ITO, H., GAUBERT, H., BUCHER, E., MIROUZE, M., VAILLANT, I. & PASZKOWSKI, J. 2011. An siRNA

pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature*, **472**, 115–120, 10.1038/nature09861.

JEDDELOH, J.A., STOKES, T.L. & RICHARDS, E.J. 1999. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nature Genetics*, **22**, 94–97, 10.1038/8803.

JI, J., TANG, D., WANG, M., LI, Y., ZHANG, L., WANG, K., LI, M. & CHENG, Z. 2013. MRE11 is required for homologous synapsis and DSB processing in rice meiosis. *Chromosoma*, **122**, 363–376, 10.1007/s00412-013-0421-1.

JOHN INNES CENTRE. 2020. Germplasm Resource Unit (GRU) | John Innes Centre Available at: https://www.jic.ac.uk/research-impact/germplasm-resource-unit/ [Accessed August 28, 2020].

JOHNSON, L.M., DU, J., HALE, C.J., BISCHOF, S., FENG, S., CHODAVARAPU, R.K., ZHONG, X., et al. 2014. SRA-and SET-domain-containing proteins link RNA polymerase v occupancy to DNA methylation. *Nature*, **507**, 124–128, 10.1038/nature12931.

JONES, G.H. & FRANKLIN, F.C.H. 2006. Meiotic Crossing-over: Obligation and Interference. *Cell*, **126**, 246–248, 10.1016/j.cell.2006.07.010.

JØRGENSEN, I.H. 1992. Discovery, characterization and exploitation of Mlo powdery mildew resistance in barley. *Euphytica*, **63**, 141–152, 10.1007/BF00023919.

KANKEL, M.W., RAMSEY, D.E., STOKES, T.L., FLOWERS, S.K., HAAG, J.R., JEDDELOH, J.A., RIDDLE, N.C., VERBSKY, M.L. & RICHARDS, E.J. 2003. Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics*, **163**, 1109–1122, 10.1093/genetics/163.3.1109.

KANNO, T., BUCHER, E., DAXINGER, L., HUETTEL, B., BÖHMDORFER, G., GREGOR, W., KREIL, D.P., MATZKE, M. & MATZKE, A.J.M. 2008. A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nature Genetics*, **40**, 670–675, 10.1038/ng.119.

KANNO, T., METTE, M.F., KREIL, D.P., AUFSATZ, W., MATZKE, M. & MATZKE, A.J.M. 2004. Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Current Biology*, **14**, 801–805, 10.1016/j.cub.2004.04.037.

KAPUSTIN, Y., SOUVOROV, A., TATUSOVA, T. & LIPMAN, D. 2008. Splign: Algorithms for computing spliced alignments with identification of paralogs. *Biology Direct*, **3**, 20,

10.1186/1745-6150-3-20.

KATO, M., MIURA, A., BENDER, J., JACOBSEN, S.E. & KAKUTANI, T. 2003. Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. *Current Biology*, **13**, 421–426, 10.1016/S0960-9822(03)00106-4.

KAWAKATSU, T., HUANG, S.C., JUPE, F., SASAKI, E., SCHMITZ, R.J., URICH, M.A., CASTANON, R., NERY, J.R., BARRAGAN, C. & HE, Y. 2016. Epigenomic diversity in a global collection of Arabidopsis thaliana accessions. *Cell*, **166**, 492–505.

KEENEY, S., GIROUX, C.N. & KLECKNER, N. 1997. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell*, **88**, 375–384, 10.1016/S0092-8674(00)81876-0.

KEENEY, S., LANGE, J. & MOHIBULLAH, N. 2014. Self-organization of meiotic recombination initiation: General principles and molecular pathways. *Annual Review of Genetics*, **48**, 187–214, 10.1146/annurev-genet-120213-092304.

KHULAN, B., THOMPSON, R.F., YE, K., FAZZARI, M.J., SUZUKI, M., STASIEK, E., FIGUEROA, M.E., et al. 2006. Comparative isoschizomer profiling of cytosine methylation: The HELP assay. *Genome Research*, **16**, 1046–1055, 10.1101/gr.5273806.

KLECKNER, N. 2006. Chiasma formation: Chromatin/axis interplay and the role(s) of the synaptonemal complex. *Chromosoma*, **115**, 175–194, 10.1007/s00412-006-0055-7.

KNÜPFFER, H., TERENTYEVA, I., HAMMER, K., KOVALEVA, O. & SATO, K. 2003. *Chapter 4 Ecogeographical diversity-a Vavilovian appraoch*. Developmen. von Bothmer, R., van Hintum, T., Knupffer, H. & Sato, K., *eds.* Elsevier, 53–76 pp, 10.1016/S0168-7972(03)80006-3.

KOSAMBI, D.D. 1944. The estimation of map distance. *Ann. Eugenics*, **12**, 505–525.

KOSZUL, R. & KLECKNER, N. 2009. Dynamic chromosome movements during meiosis: a way to eliminate unwanted connections? *Trends in Cell Biology*, **19**, 716–724, 10.1016/j.tcb.2009.09.007.

KUCZYŃSKA, A., SURMA, M., ADAMSKI, T., MIKOŁAJCZAK, K., KRYSTKOWIAK, K. & OGRODOWICZ, P. 2013. Effects of the semi-dwarfing sdw1/denso gene in barley. *Journal of Applied Genetics*, **54**, 381–390, 10.1007/s13353-013-0165-x.

KUMAR, S., STECHER, G. & TAMURA, K. 2016. MEGA7: Molecular Evolutionary Genetics

Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution*, **33**, 1870–1874, 10.1093/molbev/msw054.

KURDYUKOV, S. & BULLOCK, M. 2016. DNA methylation analysis: Choosing the right method. *Biology*, **5**, 10.3390/biology5010003.

KUSCH, S. & PANSTRUGA, R. 2017. Mlo-based resistance: An apparently universal "weapon" to defeat powdery mildew disease. *Molecular Plant-Microbe Interactions*, **30**, 179–189, 10.1094/MPMI-12-16-0255-CR.

LA, H., DING, B., MISHRA, G.P., ZHOU, B., YANG, H., BELLIZZI, M.D.R., CHEN, S., et al. 2011. A 5-methylcytosine DNA glycosylase/lyase demethylates the retrotransposon Tos17 and promotes its transposition in rice. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 15498–15503, 10.1073/pnas.1112704108.

LABABIDI, S., MEJLHEDE, N., RASMUSSEN, S.K., BACKES, G., AL-SAID, W., BAUM, M. & JAHOOR, A. 2009. Identification of barley mutants in the cultivar "lux" at the dhn loci through tilling. *Plant Breeding*, **128**, 332–336, 10.1111/j.1439-0523.2009.01640.x.

LAI, K.S., KAOTHIEN-NAKAYAMA, P., IWANO, M. & TAKAYAMA, S. 2012. A TILLING resource for functional genomics in Arabidopsis thaliana accession C24. *Genes and Genetic Systems*, **87**, 291–297, 10.1266/ggs.87.291.

LANDER, E.S., LINTON, L.M., BIRREN, B., NUSBAUM, C., ZODY, M.C., BALDWIN, J., DEVON, K., et al. 2001. Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921, 10.1038/35057062.

LANG, Z., WANG, Y., TANG, K., TANG, D., DATSENKA, T., CHENG, J., ZHANG, Y., HANDA, A.K. & ZHU, J.K. 2017. Critical roles of DNA demethylation in the activation of ripening-induced genes and inhibition of ripening-repressed genes in tomato fruit. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, E4511–E4519, 10.1073/pnas.1705233114.

LAW, J.A., AUSIN, I., JOHNSON, L.M., VASHISHT, A.A., ZHU, J.K., WOHLSCHLEGEL, J.A. & JACOBSEN, S.E. 2010. A Protein Complex Required for Polymerase V Transcripts and RNA-Directed DNA Methylation in Arabidopsis. *Current Biology*, **20**, 951–956, 10.1016/j.cub.2010.03.062.

LAW, J.A. & JACOBSEN, S.E. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, **11**, 204–220,

10.1038/nrg2719.

LEI, M., ZHANG, H., JULIAN, R., TANG, K., XIE, S. & ZHU, J.K. 2015. Regulatory link between DNA methylation and active demethylation in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 3553–3557, 10.1073/pnas.1502279112.

LEI, Y., HUANG, Y.H. & GOODELL, M.A. 2018. DNA methylation and de-methylation using hybrid site-targeting proteins. *Genome Biology*, **19**, 1DUMMY, 10.1186/s13059-018-1566-2.

LI, P., DEMIRCI, F., MAHALINGAM, G., DEMIRCI, C., NAKANO, M. & MEYERS, B.C. 2013. An Integrated Workflow for DNA Methylation Analysis. *Journal of Genetics and Genomics*, **40**, 249–260, 10.1016/j.jgg.2013.03.010.

LI, QING, GENT, J.I., ZYNDA, G., SONG, J., MAKAREVITCH, I., HIRSCH, C.D., HIRSCH, C.N., et al. 2015. RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 14728–14733, 10.1073/pnas.1514680112.

LI, X., ZHU, J., HU, F., GE, S., YE, M., XIANG, H., ZHANG, G., et al. 2012. Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. *BMC Genomics*, **13**, 300, 10.1186/1471-2164-13-300.

LI, YAN, DUAN, C.G., ZHU, X., QIAN, W. & ZHU, J.K. 2015. A DNA ligase required for active DNA demethylation and genomic imprinting in Arabidopsis. *Cell Research*, **25**, 757–760, 10.1038/cr.2015.45.

LIEBE, B., ALSHEIMER, M., HÖÖG, C., BENAVENTE, R. & SCHERTHAN, H. 2004. Telomere Attachment, Meiotic Chromosome Condensation, Pairing, and Bouquet Stage Duration Are Modified in Spermatocytes Lacking Axial Elements. *Molecular Biology of the Cell*, **15**, 827–837, 10.1091/mbc.E03-07-0524.

LINDROTH, A.M., CAO, X., JACKSON, J.P., ZILBERMAN, D., MCCALLUM, C.M., HENIKOFF, S. & JACOBSEN, S.E. 2001. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science*, **292**, 2077–2080, 10.1126/science.1059745.

LIU, R., HOW-KIT, A., STAMMITTI, L., TEYSSIER, E., ROLIN, D., MORTAIN-BERTRAND, A., HALLE, S., et al. 2015. A DEMETER-like DNA demethylase governs tomato fruit ripening.

Proceedings of the National Academy of Sciences of the United States of America, **112**, 10804–10809, 10.1073/pnas.1503362112.

LORENZ, A., MEHATS, A., OSMAN, F. & WHITBY, M.C. 2014. Rad51/Dmc1 paralogs and mediators oppose DNA helicases to limit hybrid DNA formation and promote crossovers during meiotic recombination. *Nucleic Acids Research*, **42**, 13723–13735, 10.1093/nar/gku1219.

LYNGKJAER, M.F. & CARVER, T.L.W. 2000. Conditioning of cellular defence responses to powdery mildew in cereal leaves by prior attack. *Molecular Plant Pathology*, **1**, 41–49, 10.1046/j.1364-3703.2000.00006.x.

MA, X., WANG, Q., WANG, Y., MA, J., WU, N., NI, S., LUO, T., et al. 2016. Chromosome aberrations induced by zebularine in triticale. *Genome*, **59**, 485–492, 10.1139/gen-2016-0047.

MACQUEEN, A.J. & ROEDER, G.S. 2009. Fpr3 and Zip3 Ensure that Initiation of Meiotic Recombination Precedes Chromosome Synapsis in Budding Yeast. *Current Biology*, **19**, 1519–1526, 10.1016/j.cub.2009.08.048.

MADEIRA, F., PARK, Y.M., LEE, J., BUSO, N., GUR, T., MADHUSOODANAN, N., BASUTKAR, P., et al. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, **47**, W636–W641, 10.1093/nar/gkz268.

MAHESH, S., BETHOJU, K., NALLI, A., FRANK, K. & SIDDIQI, I. 2021. Functional analysis of a conserved domain in SWITCH1 reveals a role in commitment to female meiocyte differentiation in Arabidopsis. *Biochemical and Biophysical Research Communications*, **551**, 121–126, 10.1016/j.bbrc.2021.03.007.

MAJEROVÁ, E., FOJTOVÁ, M., MOZGOVÁ, I., BITTOVÁ, M. & FAJKUS, J. 2011. Hypomethylating drugs efficiently decrease cytosine methylation in telomeric DNA and activate telomerase without affecting telomere lengths in tobacco cells. *Plant Molecular Biology*, **77**, 371–380, 10.1007/s11103-011-9816-7.

MALINOWSKA, M., NAGY, I., WAGEMAKER, C.A.M., RUUD, A.K., SVANE, S.F., THORUP-KRISTENSEN, K., JENSEN, C.S., et al. 2020. The cytosine methylation landscape of spring barley revealed by a new reduced representation bisulfite sequencing pipeline, WellMeth. *Plant Genome*, **13**, 10.1002/tpg2.20049.

MARTÍN, B., RAMIRO, M., MARTÍNEZ-ZAPATER, J.M. & ALONSO-BLANCO, C. 2009. A high-density

collection of EMS-induced mutations for TILLING in Landsberg erecta genetic background of Arabidopsis. *BMC Plant Biology*, **9**, 10.1186/1471-2229-9-147.

MARTINI, E., DIAZ, R.L., HUNTER, N. & KEENEY, S. 2006. Crossover Homeostasis in Yeast Meiosis. *Cell*, **126**, 285–295, 10.1016/j.cell.2006.05.044.

MATHIEU, O., REINDERS, J., ČAIKOVSKI, M., SMATHAJITT, C. & PASZKOWSKI, J. 2007. Transgenerational Stability of the Arabidopsis Epigenome Is Coordinated by CG Methylation. *Cell*, **130**, 851–862, 10.1016/j.cell.2007.07.007.

MATYSZCZAK, I., TOMINSKA, M., ZAKHRABEKOVA, S., DOCKTER, C. & HANSSON, M. 2020. Analysis of early-flowering genes at barley chromosome 2H expands the repertoire of mutant alleles at the Mat-c locus. *Plant Cell Reports*, **39**, 47–61, 10.1007/s00299-019-02472-4.

MAYER, K.F.X., WAUGH, R., LANGRIDGE, P., CLOSE, T.J., WISE, R.P., GRANER, A., MATSUMOTO, T., et al. 2012. A physical, genetic and functional sequence assembly of the barley genome. *Nature*, **491**, 711–716, 10.1038/nature11543.

MEISSNER, A., GNIRKE, A., BELL, G.W., RAMSAHOYE, B., LANDER, E.S. & JAENISCH, R. 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research*, **33**, 5868–5877, 10.1093/nar/gki901.

MELAMED-BESSUDO, C. & LEVY, A.A. 2012. Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, E981–E988, 10.1073/pnas.1120742109.

MERCIER, R., MEZARD, C., JENCZEWSKI, E., MACAISNE, N. & GRELON, M. 2015. The molecular biology of meiosis in plants. *Annual Review of Plant Biology*, **66**, 297–327, 10.1146/annurev-arplant-050213-035923.

MIEDANER, T. & KORZUN, V. 2012. Marker-assisted selection for disease resistance in wheat and barley breeding. *Phytopathology*, **102**, 560–566, 10.1094/PHYTO-05-11-0157.

MILNE, I., SHAW, P., STEPHEN, G., BAYER, M., CARDLE, L., THOMAS, W.T.B., FLAVELL, A.J. & MARSHALL, D. 2010. Flapjack-graphical genotype visualization. *Bioinformatics*, **26**, 3133–3134, 10.1093/bioinformatics/btq580.

MILNER, S.G., JOST, M., TAKETA, S., MAZÓN, E.R., HIMMELBACH, A., OPPERMANN, M., WEISE, S., et

al. 2019. Genebank genomics highlights the diversity of a global barley collection. *Nature Genetics*, **51**, 319–326, 10.1038/s41588-018-0266-x.

MIROUZE, M., LIEBERMAN-LAZAROVICH, M., AVERSANO, R., BUCHER, E., NICOLET, J., REINDERS, J. & PASZKOWSKI, J. 2012. Loss of DNA methylation affects the recombination landscape in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 5880–5885, 10.1073/pnas.1120841109.

MIROUZE, M., REINDERS, J., BUCHER, E., NISHIMURA, T., SCHNEEBERGER, K., OSSOWSKI, S., CAO, J., WEIGEL, D., PASZKOWSKI, J. & MATHIEU, O. 2009. Selective epigenetic control of retrotransposition in Arabidopsis. *Nature*, **461**, 427–430, 10.1038/nature08328.

MITTMANN, S.U. 2017. *Discovery of a novel meiotic E3 ubiquitin ligase by characterisation of the barley desynaptic mutant des12.w*.

MOHAMMED, J., SELESHI, S., NEGA, F. & LEE, M. 2016. Revisit to Ethiopian traditional barley-based food. *Journal of Ethnic Foods*, **3**, 135–141, 10.1016/j.jef.2016.06.001.

MONAT, C., PADMARASU, S., LUX, T., WICKER, T., GUNDLACH, H., HIMMELBACH, A., ENS, J., et al. 2019. TRITEX: chromosome-scale sequence assembly of Triticeae genomes with open-source tools. *Genome Biology*, **20**, 284, 10.1186/s13059-019-1899-5.

MORALES-RUIZ, T., ORTEGA-GALISTEO, A.P., PONFERRADA-MARÍN, M.I., MARTÍNEZ-MACÍAS, M.I., ARIZA, R.R. & ROLDÁN-ARJONA, T. 2006. DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 6853–6858, 10.1073/pnas.0601109103.

MORRIS, J.A. & HEDLEY, P.E. 2019. Microarrays for high-throughput gene expression analysis of barley. *In Methods in Molecular Biology*. 181–194., 10.1007/978-1-4939-8944-7_12.

ORTA, M.L., PASTOR, N., BURGOS-MORÓN, E., DOMÍNGUEZ, I., CALDERÓN-MONTAÑO, J.M., HUERTAS CASTAÑO, C., LÓPEZ-LÁZARO, M., HELLEDAY, T. & MATEOS, S. 2017. Zebularine induces replication-dependent double-strand breaks which are preferentially repaired by homologous recombination. *DNA Repair*, **57**, 116–124, 10.1016/j.dnarep.2017.07.002.

ORTEGA-GALISTEO, A.P., MORALES-RUIZ, T., ARIZA, R.R. & ROLDÁN-ARJONA, T. 2008. Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution

of DNA methylation marks. *Plant Molecular Biology*, **67**, 671–681, 10.1007/s11103-008-9346-0.

OSMAN, K., HIGGINS, J.D., SANCHEZ-MORAN, E., ARMSTRONG, S.J. & FRANKLIN, F.C.H. 2011. Pathways to meiotic recombination in Arabidopsis thaliana. *New Phytologist*, **190**, 523–544, 10.1111/j.1469-8137.2011.03665.x.

PATEL, D.S., MISENKO, S.M., HER, J. & BUNTING, S.F. 2017. BLM helicase regulates DNA repair by counteracting RAD51 loading at DNA double-strand break sites. *Journal of Cell Biology*, **216**, 3521–3534, 10.1083/jcb.201703144.

PATRO, R., DUGGAL, G., LOVE, M.I., IRIZARRY, R.A. & KINGSFORD, C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, **14**, 417–419, 10.1038/nmeth.4197.

PECINKA, A., ROSA, M., SCHIKORA, A., BERLINGER, M., HIRT, H., LUSCHNIG, C. & SCHEID, O.M. 2009. Transgenerational stress memory is not a general response in Arabidopsis. *PLoS ONE*, **4**, e5202, 10.1371/journal.pone.0005202.

PHILLIPS, D., JENKINS, G., MACAULAY, M., NIBAU, C., WNETRZAK, J., FALLDING, D., COLAS, I., OAKEY, H., WAUGH, R. & RAMSAY, L. 2015. The effect of temperature on the male and female recombination landscape of barley. *New Phytologist*, **208**, 421–429, 10.1111/nph.13548.

PHILLIPS, D., NIBAU, C., WNETRZAK, J. & JENKINS, G. 2012. High resolution analysis of meiotic chromosome structure and behaviour in barley (Hordeum vulgare L.) Lichten, M., *ed. PLoS ONE*, **7**, e39539, 10.1371/journal.pone.0039539.

PHILLIPS, D., WNETRZAK, J., NIBAU, C., BARAKATE, A., RAMSAY, L., WRIGHT, F., HIGGINS, J.D., PERRY, R.M. & JENKINS, G. 2013. Quantitative high resolution mapping of HvMLH3 foci in barley pachytene nuclei reveals a strong distal bias. *Journal of Experimental Botany*, **64**, 2139–2154, 10.1093/jxb/ert079.

PIFFANELLI, P., RAMSAY, L., WAUGH, R., BENABDELMOUNA, A., D'HONT, A., HOLLRICHER, K., JØRGENSEN, J.H., SCHULZE-LEFERT, P. & PANSTRUGA, R. 2004. A barley cultivation-associated polymorphism conveys resistance to powdery mildew. *Nature*, **430**, 887–891, 10.1038/nature02781.

PIKAARD, C.S., HAAG, J.R., PONTES, O.M.F., BLEVINS, T. & COCKLIN, R. 2012. A transcription fork model for pol IV and pol V-dependent RNA-directed DNA methylation. *Cold Spring*

Harbor Symposia on Quantitative Biology, **77**, 205–212, 10.1101/sqb.2013.77.014803.

POURSAREBANI, N., TRAUTEWIG, C., MELZER, M., NUSSBAUMER, T., LUNDQVIST, U., RUTTEN, T., SCHMUTZER, T., et al. 2020. COMPOSITUM 1 contributes to the architectural simplification of barley inflorescence via meristem identity signals. *Nature Communications*, **11**, 1–16, 10.1038/s41467-020-18890-y.

QIAN, W., MIKI, D., ZHANG, H., LIU, Y., ZHANG, X., TANG, K., KAN, Y., et al. 2012. A histone acetyltransferase regulates active DNA demethylation in Arabidopsis. *Science*, **336**, 1445–1448, 10.1126/science.1219416.

QUADRANA, L. & COLOT, V. 2016. Plant Transgenerational Epigenetics. *Annual Review of Genetics*, **50**, 467–491, 10.1146/annurev-genet-120215-035254.

RAZ, A., DAHAN-MEIR, T., MELAMED-BESSUDO, C., LESHKOWITZ, D. & LEVY, A.A. 2021. Redistribution of Meiotic Crossovers Along Wheat Chromosomes by Virus-Induced Gene Silencing. *Frontiers in Plant Science*, **11**, 10.3389/fpls.2020.635139.

REVENKOVA, E., EIJPE, M., HEYTING, C., HODGES, C.A., HUNT, P.A., LIEBE, B., SCHERTHAN, H. & JESSBERGER, R. 2004. Cohesin SMC1β is required for meiotic chromosome dynamics, sister chromatid cohesion and DNA recombination. *Nature Cell Biology*, **6**, 555–562, 10.1038/ncb1135.

RISSO, D., NGAI, J., SPEED, T.P. & DUDOIT, S. 2014. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*, **32**, 896–902, 10.1038/nbt.2931.

ROCHA, P.S.C.F., SHEIKH, M., MELCHIORRE, R., FAGARD, M., BOUTET, S., LOACH, R., MOFFATT, B., WAGNER, C., VAUCHERET, H. & FURNER, I. 2005. The arabidopsis HOMOLOGY-DEPENDENT GENE SILENCING1 gene codes for an S-adenosyl-L-homocysteine hydrolase required for DNA methylation-dependent gene silencing. *Plant Cell*, **17**, 404–417, 10.1105/tpc.104.028332.

RONCERET, A. & PAWLOWSKI, W.P. 2010. Chromosome dynamics in meiotic prophase I in plants. *Cytogenetic and Genome Research*, **129**, 173–183, 10.1159/000313656.

ROWLEY, M.J., ROTHI, M.H., BÖHMDORFER, G., KUCIŃSKI, J. & WIERZBICKI, A.T. 2017. Long-range control of gene expression via RNA-directed DNA methylation Mittelsten Scheid, O., *ed. PLoS Genetics*, **13**, e1006749, 10.1371/journal.pgen.1006749.

RUSSELL, J., MASCHER, M., DAWSON, I.K., KYRIAKIDIS, S., CALIXTO, C., FREUND, F., BAYER, M., et al. 2016. Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. *Nature Genetics*, **48**, 1024–1030, 10.1038/ng.3612.

RUSSELL, J., VAN ZONNEVELD, M., DAWSON, I.K., BOOTH, A., WAUGH, R. & STEFFENSON, B. 2014. Genetic diversity and ecological niche modelling of wild barley: Refugia, large-scale post-LGM range expansion and limited mid-future climate threats? Xu, M., *ed. PLoS ONE*, **9**, e86021, 10.1371/journal.pone.0086021.

SAINTENAC, C., FAURE, S., REMAY, A., CHOULET, F., RAVEL, C., PAUX, E., BALFOURIER, F., FEUILLET, C. & SOURDILLE, P. 2011. Variation in crossover rates across a 3-Mb contig of bread wheat (Triticum aestivum) reveals the presence of a meiotic recombination hotspot. *Chromosoma*, **120**, 185–198, 10.1007/s00412-010-0302-9.

SANSON, K.R., HANNA, R.E., HEGDE, M., DONOVAN, K.F., STRAND, C., SULLENDER, M.E., VAIMBERG, E.W., et al. 2018. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nature Communications*, **9**, 5416, 10.1038/s41467-018-07901-8.

SATO, S., TABATA, S., HIRAKAWA, H., ASAMIZU, E., SHIRASAWA, K., ISOBE, S., KANEKO, T., et al. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, **485**, 635–641, 10.1038/nature11119.

SAZE, H., SCHEID, O.M. & PASZKOWSKI, J. 2003. Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nature Genetics*, **34**, 65–69, 10.1038/ng1138.

SCHNEIDER, C.A., RASBAND, W.S. & ELICEIRI, K.W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, **9**, 671–675, 10.1038/nmeth.2089.

SCHREIBER, M., BARAKATE, A., UZREK, N., MACAULAY, M., SOURDILLE, A., MORRIS, J., HEDLEY, P.E., RAMSAY, L. & WAUGH, R. 2019. A highly mutagenised barley (cv. Golden Promise) TILLING population coupled with strategies for screening-by-sequencing. *Plant Methods*, **15**, 10.1186/s13007-019-0486-9.

SCHREIBER, M., MASCHER, M., WRIGHT, J., PADMARASU, S., HIMMELBACH, A., HEAVENS, D., MILNE, L., CLAVIJO, B.J., STEIN, N. & WAUGH, R. 2020. A Genome Assembly of the Barley 'Transformation Reference' Cultivar Golden Promise. *G3: Genes|Genomes|Genetics*, **10**, 1823–1827, 10.1534/g3.119.401010.

SCHWARZBACH, E. 1967. Recessive total resistance of barley to mildew (Erysiphe graminis DC f. sp. hordei Marchal) as a mutation induced by ethylmethansulfonate. *Genetika A Slechteni*, **3**, 159–162.

SHOESMITH, J.R., SOLOMON, C.U., YANG, X., WILKINSON, L.G., SHELDRICK, S., VAN EIJDEN, E., COUWENBERG, S., et al. 2021. APETALA2 functions as a temporal factor together with BLADE-ON-PETIOLE2 and MADS29 to control flower and grain development in barley. *Development*, **148**, dev194894, 10.1242/dev.194894.

SONG, L., JAMES, S.R., KAZIM, L. & KARPF, A.R. 2005. Specific method for the determination of genomic DNA methylation by liquid chromatography-electrospray ionization tandem mass spectrometry. *Analytical Chemistry*, **77**, 504–510, 10.1021/ac0489420.

STADLER, L.J. 1928. Mutations in barley induced by X-rays and radium. *Science*, **68**, 186–187, 10.1126/science.68.1756.186.

SYM, M. & ROEDER, G.S. 1995. Zip1-induced changes in synaptonemal complex structure and polycomplex assembly. *Journal of Cell Biology*, **128**, 455–466, 10.1083/jcb.128.4.455.

SZAREJKO, I., SZURMAN-ZUBRZYCKA, M., NAWROT, M., MARZEC, M., GRUSZKA, D., KUROWSKA, M., CHMIELEWSKA, B., ZBIESZCZYK, J., JELONEK, J. & MALUSZYNSKI, M. 2017. Creation of a TILLING population in barley after chemical mutagenesis with sodium azide and MNU. *In Biotechnologies for plant mutation breeding*. Springer, Cham, 91–111.

TAKUNO, S. & GAUT, B.S. 2013. Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 1797–1802, 10.1073/pnas.1215380110.

TALAMÈ, V., BOVINA, R., SANGUINETI, M.C., TUBEROSA, R., LUNDQVIST, U. & SALVI, S. 2008. TILLMore, a resource for the discovery of chemically induced mutants in barley. *Plant Biotechnology Journal*, **6**, 477–485, 10.1111/j.1467-7652.2008.00341.x.

TANG, K., LANG, Z., ZHANG, H. & ZHU, J.K. 2016. The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. *Nature Plants*, **2**, 1–10, 10.1038/nplants.2016.169.

THACKER, D., MOHIBULLAH, N., ZHU, X. & KEENEY, S. 2014. Homologue engagement controls

meiotic DNA break number and distribution. *Nature*, **510**, 241–246, 10.1038/nature13120.

THOMÄ, N.H., CZYZEWSKI, B.K., ALEXEEV, A.A., MAZIN, A. V., KOWALCZYKOWSKI, S.C. & PAVLETICH, N.P. 2005. Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54. *Nature Structural and Molecular Biology*, **12**, 350–356, 10.1038/nsmb919.

TOST, J. 2010. DNA methylation: An introduction to the biology and the disease-associated changes of a promising biomarker. *Molecular Biotechnology*, **44**, 71–81, 10.1007/s12033-009-9216-2.

TRINH, B.N., LONG, T.I. & LAIRD, P.W. 2001. DNA methylation analysis by methylight technology. *Methods*, **25**, 456–462, 10.1006/meth.2001.1268.

TSUBOUCHI, T., ZHAO, H. & ROEDER, G.S. 2006. The Meiosis-Specific Zip4 Protein Regulates Crossover Distribution by Promoting Synaptonemal Complex Formation Together with Zip2. *Developmental Cell*, **10**, 809–819, 10.1016/j.devcel.2006.04.003.

TSUKAHARA, S., KOBAYASHI, A., KAWABE, A., MATHIEU, O., MIURA, A. & KAKUTANI, T. 2009. Bursts of retrotransposition reproduced in Arabidopsis. *Nature*, **461**, 423–426, 10.1038/nature08351.

UAUY, C., PARAISO, F., COLASUONNO, P., TRAN, R.K., TSAI, H., BERARDI, S., COMAI, L. & DUBCOVSKY, J. 2009. A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. *BMC Plant Biology*, **9**, 115, 10.1186/1471-2229-9-115.

ULLRICH, S.E. 2010. *Barley: Production, improvement, and uses*. World Agri. Ullrich, S.E., *ed.* Blackwell Publishing Ltd, 1–637 pp, 10.1002/9780470958636.

UNDERWOOD, C.J., CHOI, K., LAMBING, C., ZHAO, X., SERRA, H., BORGES, F., SIMOROWSKI, J., et al. 2018. Epigenetic activation of meiotic recombination near Arabidopsis thaliana centromeres via loss of H3K9me2 and non-CG DNA methylation. *Genome Research*, **28**, 519–531, 10.1101/gr.227116.117.

UNTERGASSER, A., NIJVEEN, H., RAO, X., BISSELING, T., GEURTS, R. & LEUNISSEN, J.A.M. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*, **35**, 71–74, 10.1093/nar/gkm306.

VANYUSHIN, B.F. & ASHAPKIN, V. V. 2011. DNA methylation in higher plants: Past, present and future. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, **1809**,

360–368, 10.1016/j.bbagrm.2011.04.006.

VON BOTHMER, R. & JACOBSEN, N. 2015. Origin, Taxonomy, and Related Species. *In Barley*. John Wiley & Sons, Ltd, 19–56., 10.2134/agronmonogr26.c2.

WAGENAAR, E.B. 1960. the Cytology of Three Hybrids Involving Hordeum Jubatum L.: the Chiasma Distributions and the Occurrence of Pseudo Ring-Bivalents in Genetically Induced Asynapsis. *Canadian Journal of Botany*, **38**, 69–85, 10.1139/b60-007.

WALKER, J., GAO, H., ZHANG, J., ALDRIDGE, B., VICKERS, M., HIGGINS, J.D. & FENG, X. 2018. Sexual-lineage-specific DNA methylation regulates meiosis in Arabidopsis. *Nature Genetics*, **50**, 130–137, 10.1038/s41588-017-0008-5.

WANG, XUTONG, HU, L., WANG, XIAOFEI, LI, N., XU, C., GONG, L. & LIU, B. 2016. DNA Methylation Affects Gene Alternative Splicing in Plants: An Example from Rice. *Molecular Plant*, **9**, 305–307, 10.1016/j.molp.2015.09.016.

WANGKUMHANG, P., CHAICHOOMPU, K., NGAMPHIW, C., RUANGRIT, U., CHANPRASERT, J., ASSAWAMAKIN, A. & TONGSIMA, S. 2007. WASP: A Web-based Allele-Specific PCR assay designing tool for detecting SNPs and mutations. *BMC Genomics*, **8**, 1–9, 10.1186/1471-2164-8-275.

WEBER, M., DAVIES, J.J., WITTIG, D., OAKELEY, E.J., HAASE, M., LAM, W.L. & SCHÜBELER, D. 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genetics*, **37**, 853–862, 10.1038/ng1598.

WHITE, E.J., COWAN, C., CANDE, W.Z. & KABACK, D.B. 2004. In vivo analysis of synaptonemal complex formation during yeast meiosis. *Genetics*, **167**, 51–63, 10.1534/genetics.167.1.51.

WILCOXON, F. 1945. Individual Comparisons by Ranking Methods, 80–83 pp. Available at: http://www.jstor.org/about/terms.html. [Accessed April 24, 2021].

WILLIAMS, B.P., PIGNATTA, D., HENIKOFF, S. & GEHRING, M. 2015. Methylation-Sensitive Expression of a DNA Demethylase Gene Serves As an Epigenetic Rheostat Mittelsten Scheid, O., *ed. PLoS Genetics*, **11**, e1005142, 10.1371/journal.pgen.1005142.

WU, G., ROSSIDIVITO, G., HU, T., BERLYAND, Y. & POETHIG, R.S. 2015. Traffic lines: New tools for genetic analysis in Arabidopsis thaliana. *Genetics*, **200**, 35–45,

10.1534/genetics.114.173435.

XUE, M., WANG, J., JIANG, L., WANG, M., WOLFE, S., PAWLOWSKI, W.P., WANG, Y. & HE, Y. 2018. The number of meiotic double-strand breaks influencecrossover distribution in arabidopsis[open]. *Plant Cell*, **30**, 2628–2638, 10.1105/tpc.18.00531.

YADEGARI, R. & DREWS, G.N. 2004. Female gametophyte development. *Plant Cell*, **16**, S133–S141, 10.1105/tpc.018192.

YANG, S., OVERLANDER, M. & FIEDLER, J. 2021. Genetic analysis of the barley variegation mutant, grandpa1.a. *BMC Plant Biology*, **21**, 134, 10.1186/s12870-021-02915-9.

YELINA, N.E., CHOI, K., CHELYSHEVA, L., MACAULAY, M., DE SNOO, B., WIJNKER, E., MILLER, N., et al. 2012. Epigenetic Remodeling of Meiotic Crossover Frequency in Arabidopsis thaliana DNA Methyltransferase Mutants. *PLoS Genetics*, **8**, 10.1371/journal.pgen.1002844.

YELINA, N.E., LAMBING, C., HARDCASTLE, T.J., ZHAO, X., SANTOS, B. & HENDERSON, I.R. 2015. DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in Arabidopsis. *Genes and Development*, **29**, 2183–2202, 10.1101/gad.270876.115.

YOO, C.B., CHENG, J.C. & JONES, P.A. 2004. Zebularine: A new drug for epigenetic therapy, 910–912 pp, 10.1042/BST0320910.

ZAKHARYEVICH, K., TANG, S., MA, Y. & HUNTER, N. 2012. Delineation of joint molecule resolution pathways in meiosis identifies a crossover-specific resolvase. *Cell*, **149**, 334–347, 10.1016/j.cell.2012.03.023.

ZEMACH, A., KIM, M.Y., HSIEH, P.H., COLEMAN-DERR, D., ESHED-WILLIAMS, L., THAO, K., HARMER, S.L. & ZILBERMAN, D. 2013. The arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell*, **153**, 193–205, 10.1016/j.cell.2013.02.033.

ZHANG, H., DENG, X., MIKI, D., CUTLER, S., LA, H., HOU, Y.J., OH, J.E. & ZHU, J.K. 2012. Sulfamethazine suppresses epigenetic silencing in Arabidopsis by impairing folate synthesis. *Plant Cell*, **24**, 1230–1241, 10.1105/tpc.112.096149.

ZHANG, H., LANG, Z. & ZHU, J.K. 2018. Dynamics and function of DNA methylation in plants. *Nature Reviews Molecular Cell Biology*, **19**, 489–506, 10.1038/s41580-018-0016-z.

ZHANG, HUIMING, TANG, K., QIAN, W., DUAN, C.G., WANG, B., ZHANG, HENG, WANG, P., et al. 2014. An Rrp6-like Protein Positively Regulates Noncoding RNA Levels and DNA Methylation in Arabidopsis. *Molecular Cell*, **54**, 418–430, 10.1016/j.molcel.2014.03.019.

ZHANG, R., CALIXTO, C.P.G., MARQUEZ, Y., VENHUIZEN, P., TZIOUTZIOU, N.A., GUO, W., SPENSLEY, M., et al. 2017. A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Research*, **45**, 5061–5073, 10.1093/nar/gkx267.

ZHANG, X., YAZAKI, J., SUNDARESAN, A., COKUS, S., CHAN, S.W.L., CHEN, H., HENDERSON, I.R., et al. 2006. Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. *Cell*, **126**, 1189–1201, 10.1016/j.cell.2006.08.003.

ZHANG, X.H., YU, X.Z. & YUE, D.M. 2016. Phytotoxicity of dimethyl sulfoxide (DMSO) to rice seedlings. *International Journal of Environmental Science and Technology*, **13**, 607–614, 10.1007/s13762-015-0899-6.

ZHANG, Y., ROHDE, C., TIERLING, S., STAMERJOHANNS, H., REINHARDT, R., WALTER, J. & JELTSCH, A. 2009. DNA methylation analysis by bisulfite conversion, cloning, and sequencing of individual clones. *Methods in molecular biology (Clifton, N.J.)*, **507**, 177–187, 10.1007/978-1-59745-522-0_14.

ZHENG, B., WANG, Z., LI, S., YU, B., LIU, J.Y. & CHEN, X. 2009. Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. *Genes and Development*, **23**, 2850–2860, 10.1101/gad.1868009.

ZHONG, J., VAN ESSE, G.W., BI, X., LAN, T., WALLA, A., SANG, Q., FRANZEN, R. & VON KORFF, M. 2021. INTERMEDIUM-M encodes an HvAP2L-H5 ortholog and is required for inflorescence indeterminacy and spikelet determinacy in barley. *Proceedings of the National Academy of Sciences of the United States of America*, **118**, 10.1073/pnas.2011779118.

ZHONG, X., DU, J., HALE, C.J., GALLEGO-BARTOLOME, J., FENG, S., VASHISHT, A.A., CHORY, J., WOHLSCHLEGEL, J.A., PATEL, D.J. & JACOBSEN, S.E. 2014. Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell*, **157**, 1050–1060, 10.1016/j.cell.2014.03.056.

ZHOU, G., PARFETT, C., CUMMINGS-LORBETSKIE, C., XIAO, G.H. & DESAULNIERS, D. 2017. Two-color fluorescent cytosine extension assay for the determination of global DNA methylation. *BioTechniques*, **62**, 157–164, 10.2144/000114533.

ZHU, Y., ROWLEY, M.J., BÖHMDORFER, G. & WIERZBICKI, A.T. 2013. A SWI/SNF Chromatin-Remodeling Complex Acts in Noncoding RNA-Mediated Transcriptional Silencing. *Molecular Cell*, **49**, 298–309, 10.1016/j.molcel.2012.11.011.

ZIOLKOWSKI, P.A., BERCHOWITZ, L.E., LAMBING, C., YELINA, N.E., ZHAO, X., KELLY, K.A., CHOI, K., et al. 2015. Juxtaposition of heterozygous and homozygous regions causes reciprocal crossover remodelling via interference during Arabidopsis meiosis. *eLife*, **4**, 10.7554/eLife.03708.

8. Appendices

8.1. Primer table, use and T_M.

Primer Code	Sequence	Application	Tm (°C)	Additional comments
AT01	CTGGCAACTGGAACATTGAA	Gene Expression analysis for AT1G13440 Forward	59	Associated to TaqMan Probe #35
AT02	TTGGTGGAGAAGCTCAAGTACA	Gene Expression analysis for AT1G13440 Reverse	59	Associated to TaqMan Probe #35
AT03	GCTCTTACAGCGTTCCATCA	Gene Expression analysis for AT1G04410 Forward	59	Associated to TaqMan Probe #12
AT04	CAGTCTCCATTGCGACAAGTT	Gene Expression analysis for AT1G04410 Reverse	60	Associated to TaqMan Probe #12
AT05	GCCTATGGTAGTGGTCGGTTC	Gene Expression analysis for AT4G34620 Forward	60	Associated to TaqMan Probe #29
AT06	GGTGAAACATGTTGGCTCGT	Gene Expression analysis for AT4G34620 Reverse	60	Associated to TaqMan Probe #29
AT07	GGAGATCTAGTGAACCGTCTCAA	Gene Expression analysis for AT4G16215 Forward	59	Associated to TaqMan Probe #93
AT08	CATTGCCGGAACGATCTC	Gene Expression analysis for AT4G16215 Reverse	60	Associated to TaqMan Probe #93
AT09	AATACCGGAATTTCCATCTCC	Gene Expression analysis for AT2G17690 Forward	59	Associated to TaqMan Probe #34
AT10	TGCGAGAACGTCCTAGTGC	Gene Expression analysis for AT2G17690 Reverse	60	Associated to TaqMan Probe #34
AT11	GAGAGGAAGAGGAGACGAAGG	Gene Expression analysis for AT2G11773 Forward	59	Associated to TaqMan Probe #9
AT12	TCTCTTAATGCCTTTCCTTTGC	Gene Expression analysis for AT2G11773 Reverse	59	Associated to TaqMan Probe #9

AT13	CAACATCATTAAAGCCGCAGT	Gene Expression analysis for AT5G20630 Forward	60	Associated to TaqMan Probe #19
AT14	CCATTGATGCCTGCGTAAG	Gene Expression analysis for AT5G20630 Reverse	60	Associated to TaqMan Probe #19
AT15	ATTGAAAGGATCGGCTGAGA	Gene Expression analysis for AT2G01520 Forward	59	Associated to TaqMan Probe #25
AT16	GTCGGGGAAGAGGTGGTT	Gene Expression analysis for AT2G01520 Reverse	59	Associated to TaqMan Probe #25
AT17	TGAACGTTGGCATATTCACC	Gene Expression analysis for AT1G73330 Forward	59	Associated to TaqMan Probe #126
AT18	CATCGCTGGTCAAGGCTAAG	Gene Expression analysis for AT1G73330 Reverse	60	Associated to TaqMan Probe #126
MT01	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCACAAACTAGCAAAGAAATGAGATGGAAGG	Amplification of Met1 amplicon a, Forward	63	Abandoned, no amplification
MT02	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTGGCCCGCTCAACCTTC	Amplification of Met1 amplicon a, Reverse	63	Abandoned, no amplification
MT03	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGAGTTGGTGTCTGTCAATCTCCGA	Amplification of Met1 amplicon b, Forward	63	
MT04	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAGGCCAATGTTTCTTCCTGCCT	Amplification of Met1 amplicon b, Reverse	64	
MT05	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTCCATGACTTCTTATACATCAGGCCTG	Amplification of Met1 amplicon c, Forward	63	
MT06	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATCCACAGGCACATTAATTACATCTTCACT	Amplification of Met1 amplicon c, Reverse	63	
MT07	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGTCACCATCACTAGTTCCATGCTTG	Amplification of Met1 amplicon d, Forward	63	

MT08	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGCACCAGCTGCCTTCCGAGC	Amplification of Met1 amplicon d, Reverse	64	
MT09	TCGTCGGCAGCGTCAGATGTGTATAAGAGAC AGTCAGGCAAATGGATGGAGGTGC	Amplification of Met1 amplicon e, Forward	63	
MT10	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGTCCCCACATTTGTCCATGATTGCC	Amplification of Met1 amplicon e, Reverse	64	
MT11	TCGTCGGCAGCGTCAGATGTGTATAAGAGAC AGACTGCAATGTGATCTTGAAGTAGGTGAA	Amplification of Met1 amplicon f, Forward	63	
MT12	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGAGGAATGCTAAAATCATCTCACACTGA ACT	Amplification of Met1 amplicon f, Reverse	63	
MT13	TCGTCGGCAGCGTCAGATGTGTATAAGAGAC AGTGGTGGCCCTCCCTGTCAG	Amplification of Met1 amplicon g, Forward	64	
MT14	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGCCTGCTTCTAAGATTCCAAAACGAACCT	Amplification of Met1 amplicon g, Reverse	64	
MT15	TCGTCGGCAGCGTCAGATGTGTATAAGAGAC AGGGTTACCAGGTGCTAATTCTTGCCA	Amplification of Met1 amplicon h, Forward	63	
MT16	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGAAAACTGAATGGCCTTGGAATATTACC TCA	Amplification of Met1 amplicon h, Reverse	63	
MT17	TCGTCGGCAGCGTCAGATGTGTATAAGAGAC AGCGAAAGTGGAAAATGGTGCAAGTAAACT	Amplification of Met1 amplicon i, Forward	63	
MT18	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGGTCCACCATTTGCCCAGAAGATAGC	Amplification of Met1 amplicon i, Reverse	64	
MT19	TCGTCGGCAGCGTCAGATGTGTATAAGAGAC AGGGTGTTTGTGAGATGTTAATTCCGGGA	Amplification of Met1 amplicon j, Forward	64	
MT20	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGTGCCCGCAAACTGGTAGCTG	Amplification of Met1 amplicon j, Reverse	63	

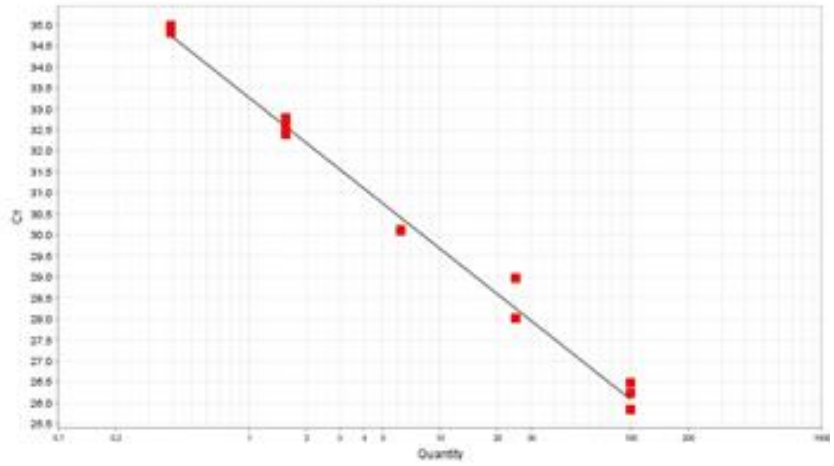
MT21	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGAGGTTAGATTGGGAGGGCA	Amplification of Met1 amplicon k, Forward	64	
MT22	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCCTTGGGAGTTCTACAGACGCA	Amplification of Met1 amplicon k, Reverse	64	
MT23	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTGGCCACCTCTCATGCC	Amplification of Met1 amplicon l, Forward	63	
MT24	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTCTGGCCCTAATTTTGTCAGGTT	Amplification of Met1 amplicon l, Reverse	63	
MT25	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGATAACCAGGATGCAGGAGTTCC	Amplification of Met1 amplicon m, Forward	62	
MT26	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGTACAGGATCACACCAAGGAGCA	Amplification of Met1 amplicon m, Reverse	63	
MT27	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGACCTGCCAGATGAGAAGGTAAAGT	Amplification of Met1 amplicon n, Forward	64	Redundant with MT19
MT28	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACACATTTGAGCCTATTTGTCATCAGATG	Amplification of Met1 amplicon n, Reverse	64	Redundant with MT20
MT29	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCAAAGAAATGAGATGGAAGGGTGAA	Amplification of Met1 amplicon a, Forward	63	Replaces MT01
MT30	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTTGGCCCGCTCAACCTTC	Amplification of Met1 amplicon a, Reverse	63	Replaces MT02
MT31	TCGCACAAGGAAAATTGGTG	Amplification of Met1 for validation of IP-153, Forward	62	
MT32	TCCTCAATACGCCCAAATCC	Amplification of Met1 for validation of IP-153, Reverse	62	
MT33	GGATTTGGGCGTATTGAGGA	Amplification of Met1 for validation of IP-174, Forward	62	

MT34	GAGCAAGAAGAACCGGCAAT	Amplification of Met1 for validation of IP-174, Reverse	62	
DT01	GCTGGCCACCTCTCATGCC	Amplification of Ddm1 for validation of IP-239, Forward	61	
DT02	AGTACAGGATCACACCAAGGAGCA	Amplification of Ddm1 for validation of IP-239, Reverse	61	
DT03	TCAGGCAAATGGATGGAGGTGC	Amplification of Ddm1 for validation of IP-176, Forward	61	
DT04	TCCCCACATTTGTCCATGATTGCC	Amplification of Ddm1 for validation of IP-176, Reverse	61	
MV01	CCAGAGGATTCACCGCCTTG	Amplification of Met1 for validation of E151V and V208L, Forward	68	
MV02	ACAGAGCGAACAACACTAGCA	Amplification of Met1 for validation of E151V and V208L, Reverse	66	
DV01	TGGAGATGACACATCATTCTGAGTGG	Amplification of Ddm1 for validation of A225E and A253D, Forward	68	
DV02	GCAGACCATTCCCTTTAAGATGAGCA	Amplification of Ddm1 for validation of A225E and A253D, Reverse	69	
DV03	TGCTCCCCTTTCCACCCTGT	Amplification of Ddm1 for validation of K356E, Forward	72	
DV04	GTATTTTTCAACCGATGCCCCTGC	Amplification of Ddm1 for validation of K356E, Reverse	69	
DV05	GACCCCTGCTTGGCCACATT	Amplification of Ddm1 for validation of D766A, Forward	71	
DV06	CAACACGACCTCCCAGCCAG	Amplification of Ddm1 for validation of D766A, Reverse	71	
CC01	GAACACCTGCCCTTTGTTGAAGGTTGAGCGG GCCAAAGGTTTTAGAGCTAGAAATAGCAAG	Amplification of double sgRNA CRISPR construct for Met1 and Ddm1, Forward	58	Includes AarI restriction sites

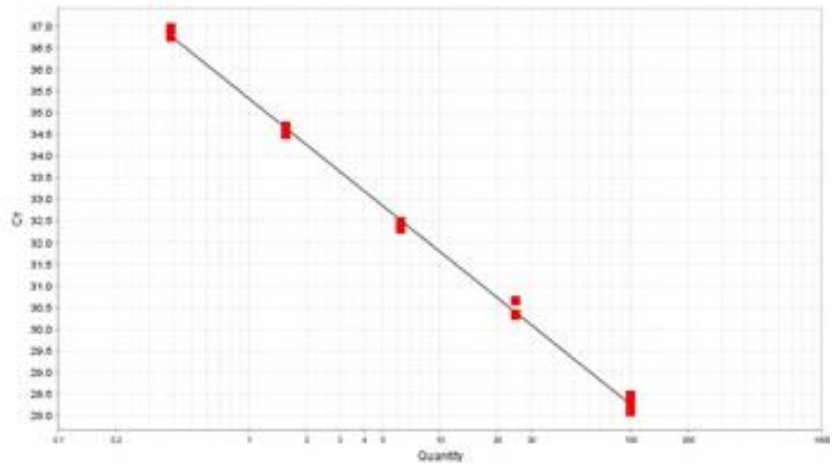
CC01	GAACACCTGCCCTTAAACCCGTTGGCCTTGA GCTCGCCAACACAAGCGACAGCGCGCGGGT	Amplification of double sgRNA CRISPR construct for Met1 and Ddm1, Reverse	71	Includes AarI restriction sites
M13-F	GTTTTCCCAGTCACGAC	Colony PCR in pGEM-T Easy, Forward	60	
M13-R	CAGGAAACAGCTATGAC	Colony PCR in pGEM-T Easy, Reverse	56	

8.2. RT-qPCR Primer Efficiency

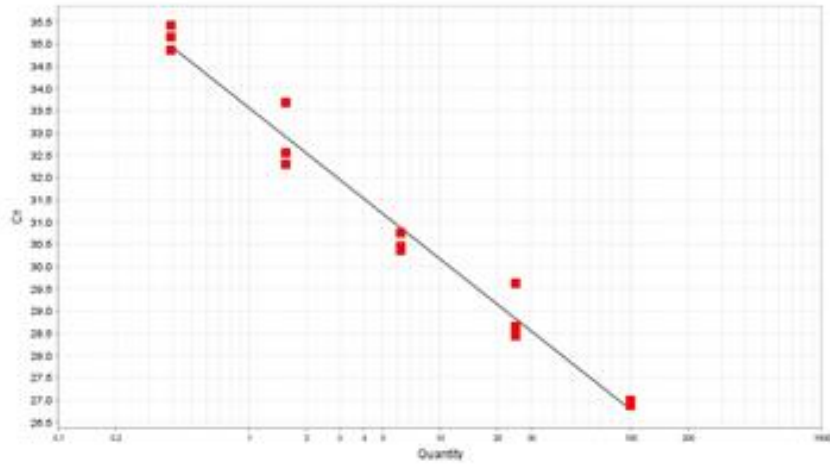
AT1G13440



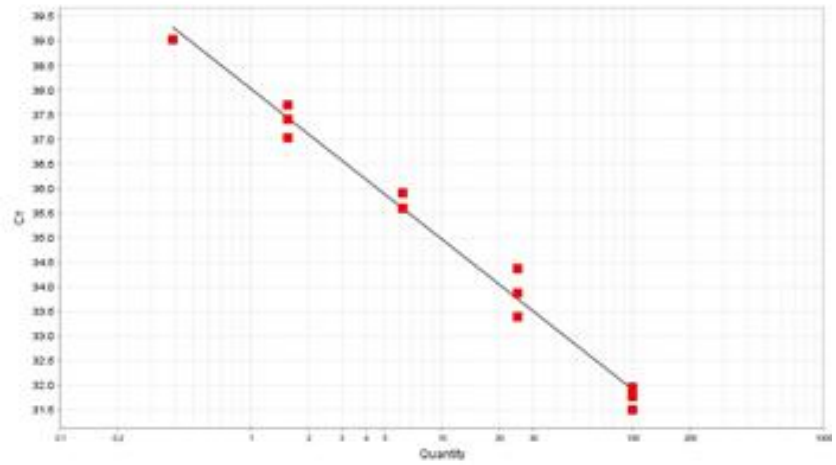
AT4G16215



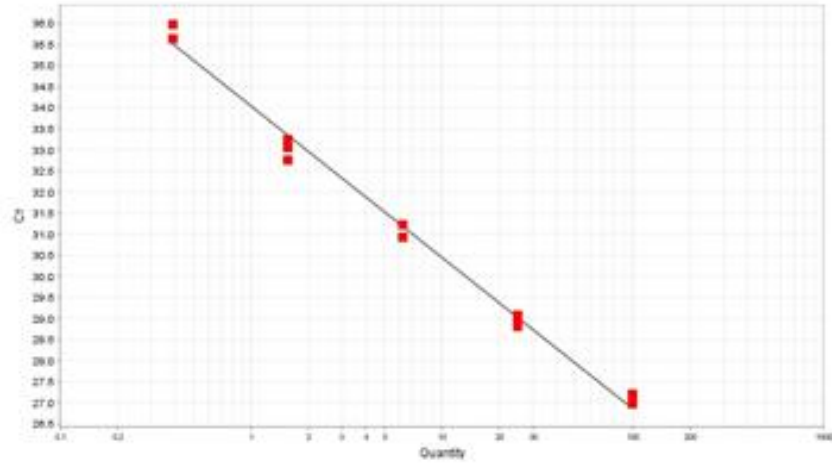
AT2G17690



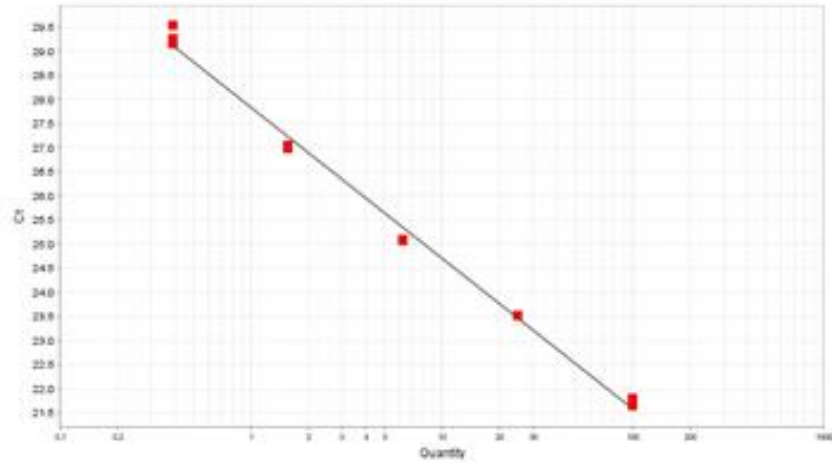
AT2G11773

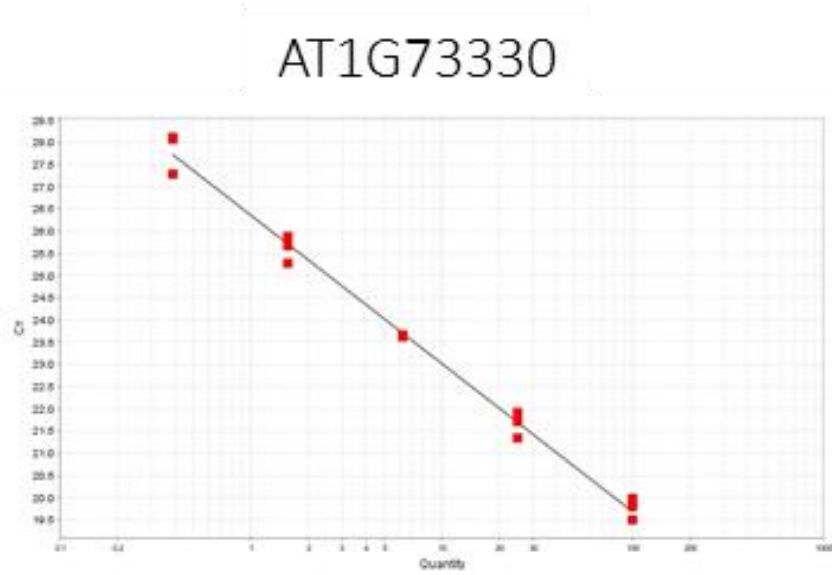


AT5G20630

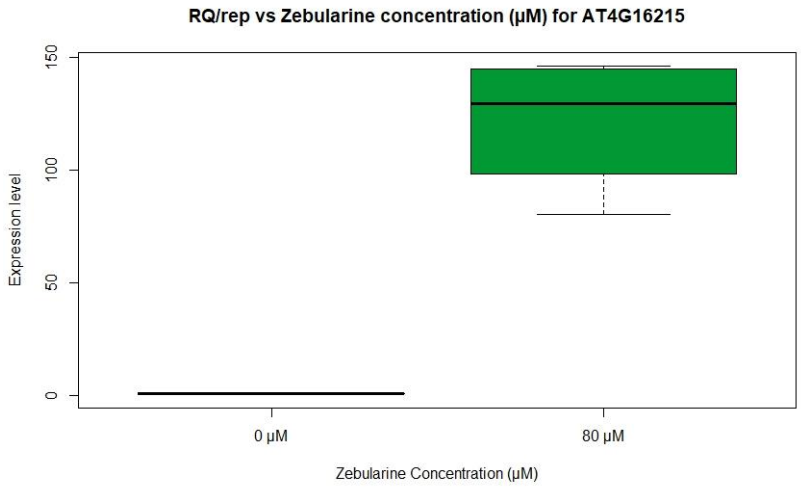


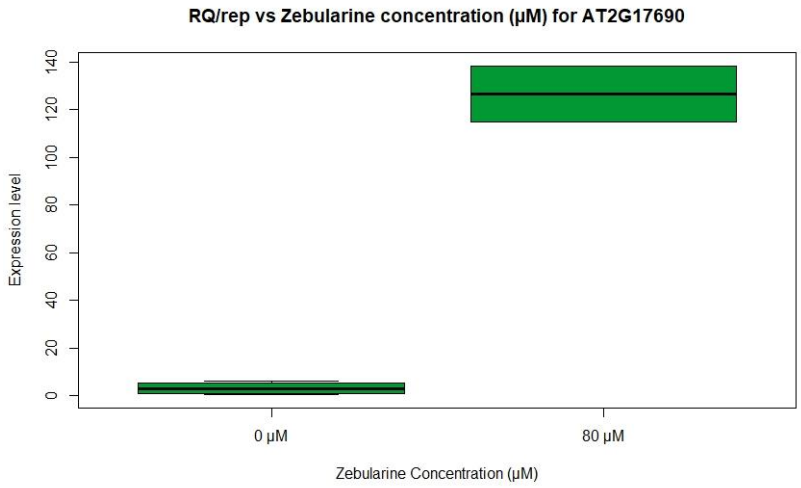
AT2G01520

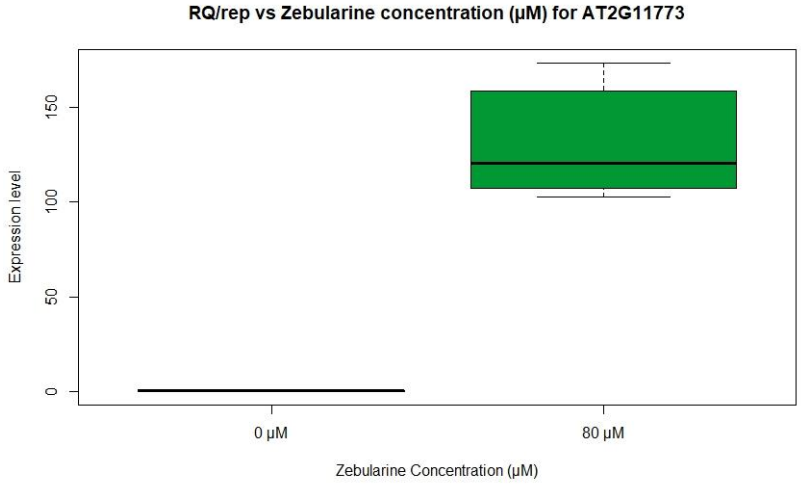


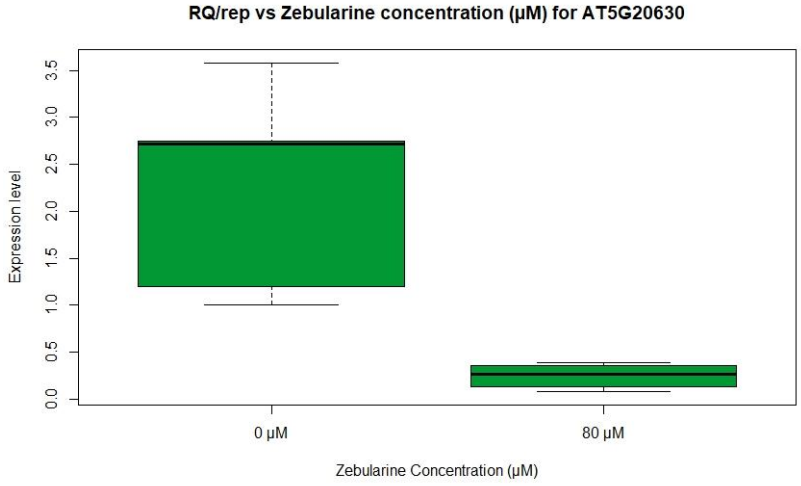


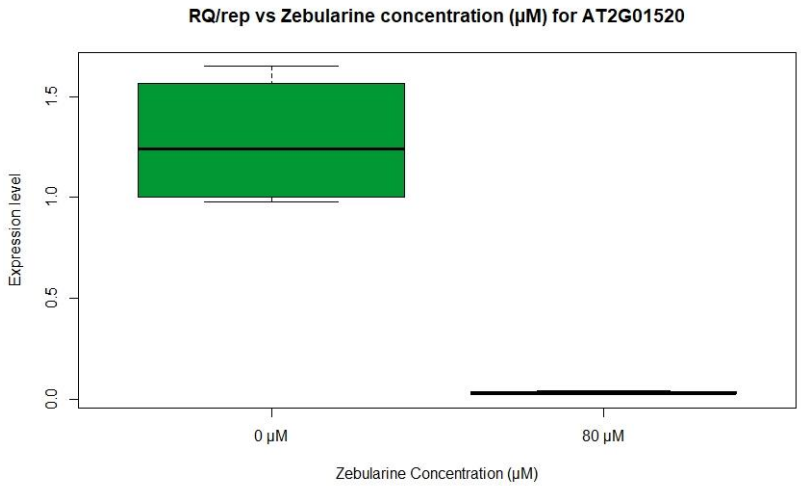
8.3. RT-qPCR Gene expression changes

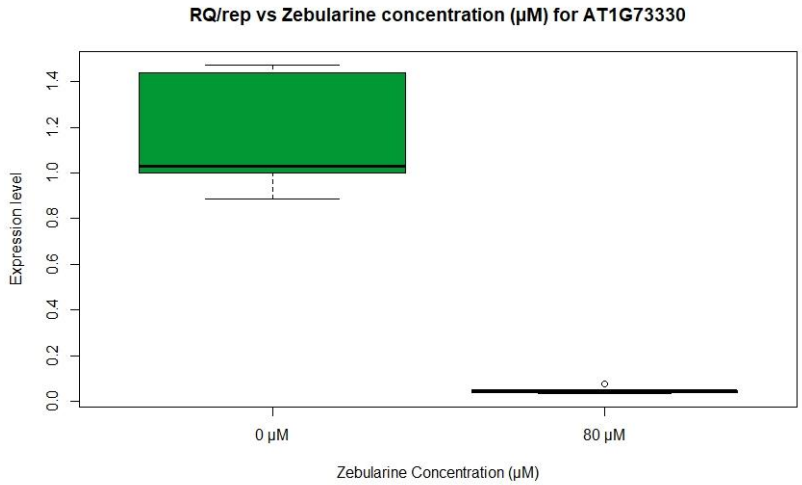





RQ/rep vs Zebularine concentration (µM) for AT2G11773
Expression level
0
50
100
150
0 µM
80 µM
Zebularine Concentration (µM)


RQ/rep vs Zebularine concentration (µM) for AT5G20630
Expression level
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0 µM
80 µM
Zebularine Concentration (µM)


RQ/rep vs Zebularine concentration (µM) for AT2G01520
Expression level
0.0
0.5
1.0
1.5
0 µM
80 µM
Zebularine Concentration (µM)



8.4. Media Recipes

8.4.1. SOB Medium

For 1 L, to 950 mL of deionized H_2O, add 20 g Tryptone, 5 g Yeast extract and 0.5 g NaCl. Shake until the solutes have dissolved. Add 10 mL of a 250 mM solution of KCl. (This solution is made by dissolving 1.86 g of KCl in 100 mL of deionized H_2O.) Adjust the pH of the medium to 7.0 with 5N NaOH (~0.2 mL). Adjust the volume of the solution to 1 L with deionized H_2O. Sterilize by autoclaving for 20 minutes at 15 psi (1.05 kg/cm^2) on liquid cycle. Just before use, add 5 mL of a sterile solution of 2M $MgCl_2$. (This solution is made by dissolving 19 g of $MgCl_2$ in 90 mL of deionized H_2O. Adjust the volume of the solution to 100 mL with deionized H_2O and sterilize by autoclaving for 20 min at 15 psi [1.05 kg/cm^2] on liquid cycle.)

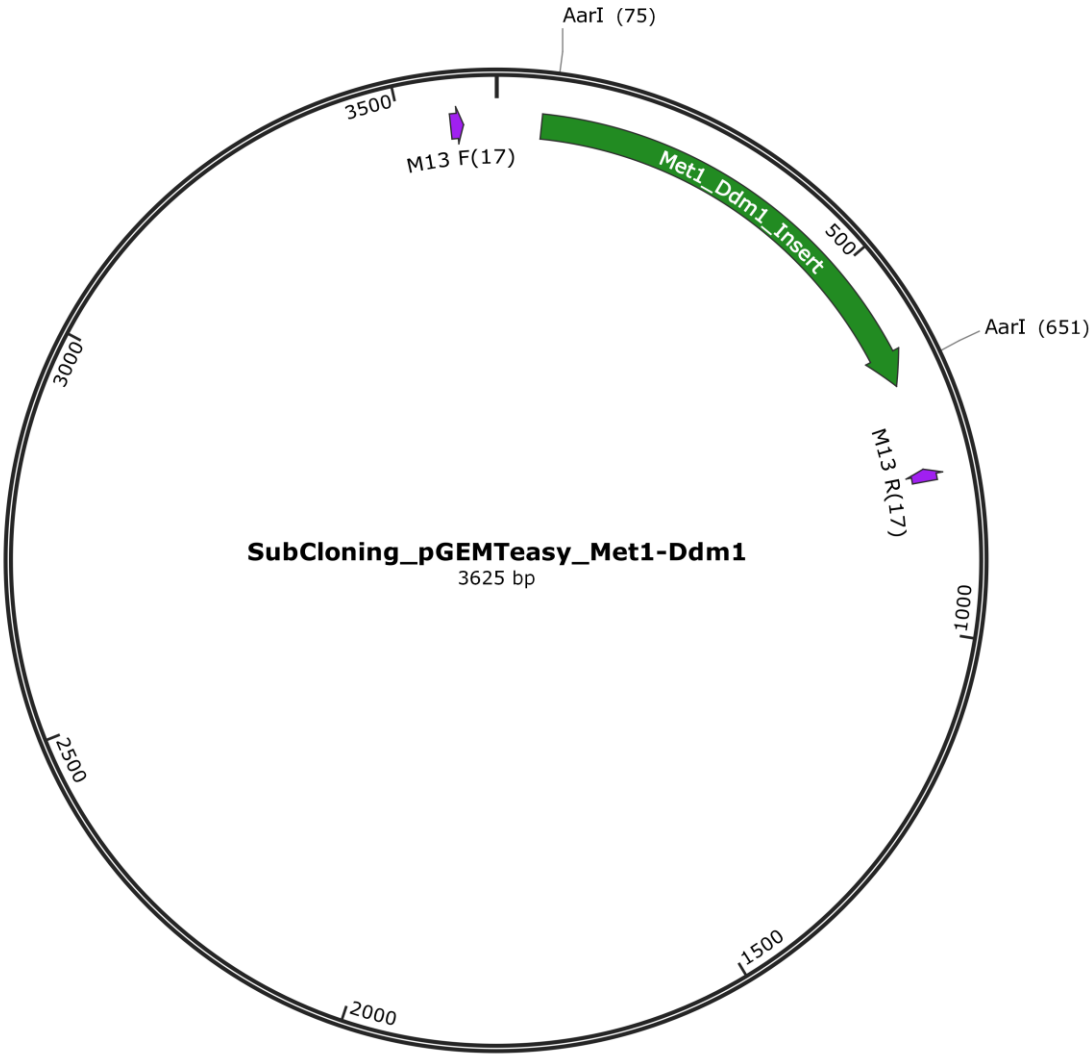
8.4.2. LB Liquid Medium

For 1 L, dissolve 10 g of Bacto-Tryptone, 5 g of Yeast extract, and 10 g of NaCl. Adjust the pH to 7.5 with NaOH. Adjust the volume of the solution to 1 L with deionized H_2O. Sterilize by autoclaving.

8.4.3. YEB medium

For 1 L, dissolve 5 g Beef extract, 1 g Yeast extract, 5 g peptone, 5 g sucrose and 0.5 g $MgCl_2$. Adjust the volume of the solution to 1 L with deionized H_2O. Sterilize by autoclaving.

8.5. Vector Map for subcloning of *Met1* and *Ddm1* sgRNA construct in pGEM-T Easy



8.6. Vector Map of binary vector pBract214m with sgRNAs for *Met1* and *Ddm1*